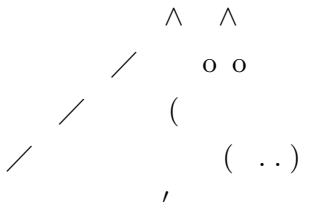
Modelling complex dependencies inherent in spatial and spatio-temporal point pattern data

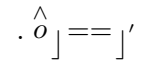
University of
St Andrews

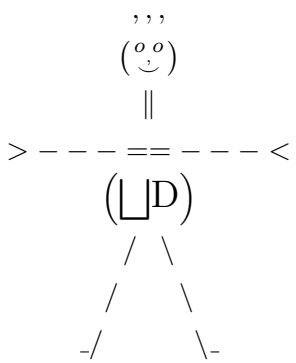
Charlotte M. Jones-Todd

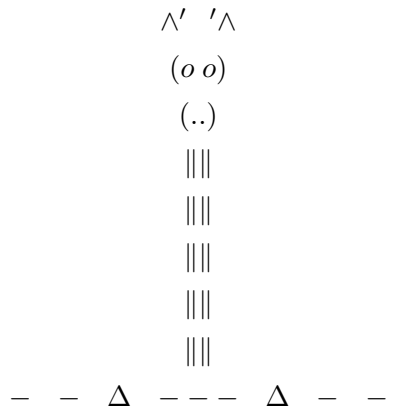
A thesis submitted for the degree of
Doctor of Philosophy
at the University of St Andrews

5th May 2017









Declaration

1. **Candidates declarations:**

   I, Charlotte Moragh Jones-Todd, hereby certify that this thesis, which is approximately 42,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

   I was admitted as a research student in September, 2103 as a candidate for the degree of PhD in Statistics; the higher study for which this is a record was carried out in the University of St Andrews between September 2013 and December 2016.

   Date: 5th May 2017 Signature of candidate:

2. **Supervisors declaration:**

   I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

   Date: 5th May 2017 Signature of supervisor:

3. **Permission for publication: (to be signed by both candidate and supervisor)**

   In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

   The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

   PRINTED COPY: Embargo on all of print copy for a period of two years on the following grounds:

   - Publication would preclude future publication.

ELECTRONIC COPY: Embargo on all of electronic copy for a period of two years on the following grounds:

- Publication would preclude future publication.

Supporting statement for electronic embargo request

The ability to publish work found within this thesis may be compromised if it is not embargoed.

Date: 5th May 2017 Signature of candidate:

Signature of supervisor:

Abstract

Point processes are mechanisms that beget point patterns. Realisations of point processes are observed in many contexts, for example, locations of stars in the sky, or locations of trees in a forest. Inferring the mechanisms that drive point processes relies on the development of models that appropriately account for the dependencies inherent in the data. Fitting models that adequately capture the complex dependency structures in either space, time, or both is often problematic. This is commonly due to—but not restricted to—the intractability of the likelihood function, or computational burden of the required numerical operations.

This thesis primarily focuses on developing point process models with some hierarchical structure, and specifically where this is a latent structure that may be considered as one of the following: (i) some unobserved construct assumed to be generating the observed structure, or (ii) some stochastic process describing the structure of the point pattern. Model fitting procedures utilised in this thesis include either (i) approximate-likelihood techniques to circumvent intractable likelihoods, (ii) stochastic partial differential equations to model continuous spatial latent structures, or (iii) improving computational speed in numerical approximations by exploiting automatic differentiation.

Moreover, this thesis extends classic point process models by considering multivariate dependencies. This is achieved through considering a general class of joint point process model, which utilise shared stochastic structures. These structures account for the dependencies inherent in multivariate point process data. These models are applied to data originating from various scientific fields; in particular, applications are considered in ecology, medicine, and geology. In addition, point process models that account for the second order behaviour of these assumed stochastic structures are also considered.

Contents

Part I

Introduction

Chapter 1

Introduction

1.1 The dimensionality of data

All data have dimensionality. In the physical world this may refer to either a spatial reference, a temporal index, or both. Proximity of data—either spatially, or temporally—often leads to similarity in the observations. This similarity, simply due to the adjacency of the data, must be accounted for when endeavouring to glean meaningful interpretations. Consider animal abundance recorded at neighbouring locations, blood sugar levels measured at proximate times, or nearby sea temperature recordings. All situations contextually present the same issue: the observed values are (spatially, temporally or spatio-temporally) correlated. Due to the dependence between the observations, many traditional statistical methods—which assume independent and identically distributed errors—are not appropriate. This thesis concentrates on the analysis of spatial and spatio-temporal data, specifically, on constructing models that account for spatial or spatio-temporal dependence. The appropriate analysis of spatial or spatio-temporal data has relevance in a range of fields, including ecology, geology, medicine and international relations. Thus, methodology is required that can characterise the inherent spatial or temporal structure, permitting one to either infer relationships, classify the data's structural arrangement, or both.

Spatial data can be classified to fall into one of three broad categories: (i) lattice data, where the observations relate to defined spatial areas (e.g., postcode districts); (ii) geostatistical data, where the observed value is assumed to be continuously spatially indexed, yet only measured at discrete locations; or (iii) point pattern data, where the spatial distribution of either objects or events is itself of interest. In the latter case, characteristics of the points provide further information that may contribute to the mechanisms governing structure of the data. Including temporal information in many cases adds to the complexity of the analysis, yet, can be invaluable in inferring either long-term or short-term trends. Again, the intrinsic autocorrelation must be accounted for, however, in some cases it may not be of particular interest.

Inferring the inherent structure of spatial or spatio-temporal data is paramount to understanding the processes which proffer the observations. This structure often represents natural phenomena, which typically exist independently of the observation at a given location. Yet, dependent on the type of spatial data considered, despite apparently similar structures, interpretations differ. For example, such a latent process might be thought of as some quantifiable continuous variable such as soil quality, or a more abstract mechanism like species interaction, which is not a physical construct. The three most common varieties of spatial data mentioned above are outlined below, along with a brief summary of typical questions posed by each data type.

1.1.1 Lattice data

Lattice data refer to data collected on some discrete lattice or grid, where there is some response associated with each discretized region. Specifically, the data locations refer to a finite number of sites that together typically constitute the entire study region. Such data are common in governmental policy, and usually refer to some economic measure of a constituency. More commonly, modelling lattice data is prevalent in the context of disease mapping (Schrödle and Held, 2011; Assunção et al., 2001; Richardson et al., 2006). Disease mapping models are frequently formulated in a hierarchical framework where the aim is to adequately describe the pattern of disease counts, and aid in the identification of regions with high incidence, usually over time. Here, the spatio-temporal latent structure can be thought of as accounting for the spatio-temporal autocorrelation between neighbouring regions at adjacent time points.

One may also commonly see lattice data in ecology, whereby the data refer to the number of plants in quadrats—a typical way to summarise point pattern data, see Section 1.1.3. Figure 1.1 (a) illustrates an example of simple lattice data, where the colours of the gridcells represent the observed values. Here the spatial index of the data refers to reference locations defined on discrete spatial elements (e.g., postcode areas or quadrats).

1.1.2 Geostatistical data

Geostatistical data usually refer to natural phenomena that vary consistently in either space or in both space and time (e.g., soil porosity, animal density, or temperature). Observations are made at a finite number of locations, either due to some sampling procedure or at predefined locations (e.g., meteorological stations). The locations of the measuring sites are typically unrelated to the scientific phenomena of interest. These observations are used to infer the spatially continuous process operating within a region, and are spatially dependent. Figure 1.1 (b) shows a typical geostatistical dataset, where the size of the plotting characters represents the magnitude of the spatially continuous process at the observed locations.

The observations can be thought of as noisy versions of the true value at a set of n known locations, $\mathbf{s}_1, \cdots, \mathbf{s}_n$. Let $Y(\mathbf{s})$ be the true value, where $\mathbf{s} \in \mathbb{R}^d$, then assuming additive measurement error, the observations (conditional on $Y(\mathbf{s})$) can be written as,

$$Z(\mathbf{s}_i) = Y(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \quad i = 1, \cdots, n, \tag{1.1}$$

where $\epsilon(\mathbf{s}_i)$ is a zero-mean white noise process with variance σ^2_ϵ.

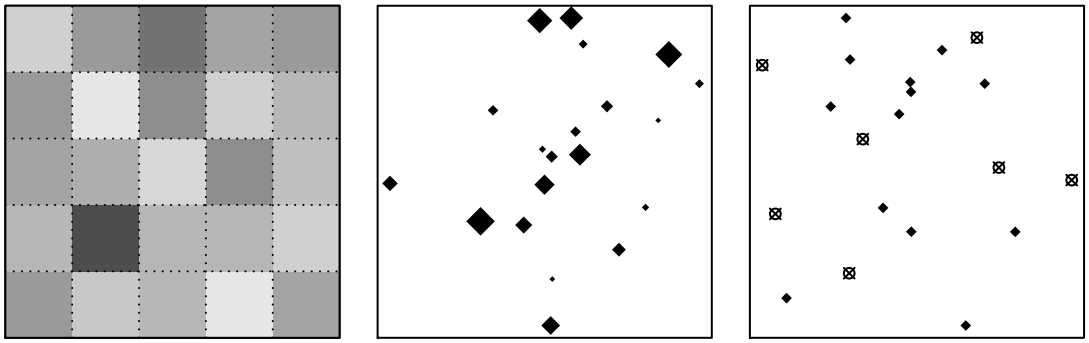
A well accepted method for the analysis of geostatistical data is Kriging, initially proposed by Matheron (1976), that aims to predict values of $Y(\cdot)$ based on the noisy data $Z(\mathbf{s}_1), \cdots, Z(\mathbf{s}_n)$. This optimal linear spatial prediction at a given location uses a weighted average of neighbouring samples. Here, the observed data are considered to be the result of some continuous random process, and as such the represented physical phenomenon can be estimated at unobserved locations, along with quantification of the associated error. For a more comprehensive account of geostatistical data and the models used to infer about the dynamics of the systems of interest, see Diggle and Ribeiro (2007) and Diggle et al. (1998).

1.1.3 Point pattern data

Point patterns are formed by the locations of objects or events in some region, $\Omega \in \mathbb{R}^d$. These points are either completely randomly distributed or exhibit some structure, such as regularity or clustering. The point pattern formed by the points is considered to be a realisation of a random variable, a point process. In comparison to either lattice data (Section 1.1.1) or geostatistical data (Section 1.1.2), the locations themselves are the response of interest. Consider the locations of trees in a forest, or the locations of recorded terrorist attacks. In each case inferring the spatial distribution of these points is paramount to understanding the processes proffering the observed structure of the points. Furthermore, characteristics of these points may be recorded. These characteristics, known as marks, may offer information as to the processes governing the spatial distribution of points. A mark is only observable at the location of a point. For example, if points are trees, then the marks may be their girths. Thus, a marked point pattern is different from geostatistical data, where measurements can be taken anywhere in continuous space. Illian et al. (2008) provide an in-depth overview of point pattern data and the statistical analysis thereof.

Figure 1.1 (c) shows a marked point pattern whereby some factor characteristic of the points is illustrated through the two different plotting symbols. Classically, a lattice type approach has been used to infer the structure of point pattern data, whereby the number of objects are counted in a finite number of gridcells overlain onto Ω. However, inference is sensitive to the choice of gridcell size: too large and any structure is lost, too small and the model overfits.

Perhaps the most commonly encountered point process is the homogeneous Poisson process. This is where (i) the number of points, in Ω, follows a Poisson distribution, and (ii) the



(a) (b) (c)

Figure 1.1: (a) Lattice data where observations, represented by different greys, refer to some finite number of spatial areas. (b) Typical geostatistical data, where the size of the points refer to the value of some observed continuous measure at the locations sampled (c) A typical point pattern generated by locations of objects or events, considering the different plotting characters leads to a marked point pattern where each plotting character indicates some characteristic of the point. In each case the data are observed in 2-dimensions.

points are independently and uniformly scattered in Ω. Point (i) above leads to the expected number of points in the process being given by $\lambda\,\nu(\cdot)$ for some constant and positive λ and volume $\nu(\cdot)$. Here λ is the characteristic parameter, called the intensity of the process. The homogeneous Poisson process can be thought of as the founding point process as it describes perhaps the most simple of point pattern. One straightforward extension of this process is to allow the intensity to vary in space, leading to an intensity function, $\lambda(s)$. The point process is now an inhomogeneous Poisson process. In later sections more complex point processes are discussed, such as point processes with latent components that define the structure of the point pattern (Section 1.2.2).

Let $N(\Omega)$ denote a point process (i.e., a random variable), in some region Ω, and let any realisation of $N(\Omega)$ be a point pattern. A point process is thus a mathematical model of a point pattern. Spatial structures inherent in many real world point pattern data are only observed in one-, two-, or three-dimensional space. Only the latter two are mainly considered in this thesis. In order to analyse the geometrical structure of point patterns, point process statistics are used; typically, these are aimed at describing the correlations between points (discussed in Section 1.2). This thesis chiefly focuses on developing methodologies which consider complex structures, and dependencies intrinsic in any $N(\Omega)$, which are assumed to represent real world phenomena.

1.2 Point process methodology

Broadly speaking, this thesis focuses on the analysis of spatial and spatio-temporal point pattern data. An overview of this thesis is provided in Section 1.4, but first it is necessary to introduce existing point process methodology, and some common notation which will be used

throughout this thesis. Functional summary characteristics are described in Section 1.2.1, and popular point process models are detailed in Section 1.2.2. One particularly flexible method which can be used to fit spatial or spatio-temporal point process models, used throughout the majority of this thesis, is summarised in Section 1.3.

Throughout this thesis, effort has been made to maintain consistent mathematical notation. However, in circumstances where this was not possible deviations are made explicit. Where possible, point process notation follows that typically used in the point process literature. As detailed above, a point pattern is a realisation of a spatial point process, denoted $N(\cdot)$, although typically herein a spatial point process is simply referred to as a point process. Thus, $N(\Omega)$ can be thought of in terms of a random variable representing the number of points of $N(\cdot)$ in Ω (defining some domain). For clarification, this thesis, unless otherwise stated, assumes $N(\cdot)$ is a stochastic process governing the locations of points in space (in this thesis only one-, two-, and three-dimensional space is considered). Another definition of a $N(\cdot)$, is one where the point process is considered to be a random counting measure; specifically for $\Omega \in \mathbb{R}^d$, $N(\Omega)$ refers to the random number of points in Ω (see Illian et al. (2008) for further details). In addition $N(\Omega)$ is assumed to be locally finite; that is, $N(\Omega) < \infty$. Therefore, $N(\cdot)$ is herein described with reference to number distributions; specifically by the univariate and multivariate distributions respectively,

$$\begin{array}{cc} \mathbf{P}(N(\Omega) = n) & \text{and} \quad \mathbf{P}(N(\Omega_1) = n_1, \cdots, N(\Omega_k) = n_k) \\ \text{for } n = 0, 1, \cdots & \text{for } n_1 = 0, 1, \cdots n_k = 0, 1, \cdots \end{array}$$

The intensity of a point process is typically denoted λ, where $\lambda(\cdot)$ denotes that the intensity is a function (typically of space, $\mathbf{s}$); however, Chapters 2 and 3 are centred around the Palm intensity function (Palm i.f), denoted $\lambda_p(\cdot)$ (a function of distance, r, and some parameters, $\boldsymbol{\theta}$). Now, $\boldsymbol{\theta}$ in all chapters refers to a vector of parameters, which are defined in each specific case.

Before moving onto introducing point process methodology, the reader should be made aware of two underlying assumptions which are made in all bar the penultimate chapter of this thesis. These assumptions relate to the point process models discussed, and are : (i) **isotropy**, where the distribution of a point process is invariant under rotations about the origin, and (ii) **stationarity**, where the distribution of a point process is invariant under translations. The following sections discuss the case when both of these assumptions are made.

1.2.1 Point process summary statistics

Functional summary characteristics are popular for inferring the geometric structure of a point pattern. Such functional characteristics offer more information than simplistic numerical characteristics (i.e., constant intensity, λ). Such functional summaries of point patterns

are typically used to ascertain a pattern's deviation from complete spatial randomness (CSR), (i.e., deviation from a homogeneous Poisson process). The most commonly utilised functional characteristics (functions of distance r) are (i) $H(\cdot)$ empty space function (spherical contact distribution function), (ii) $G(\cdot)$ nearest neighbour distribution function, and (iii) $K(\cdot)$ Ripley's K-function. A distinction should be made between these functions due to where the distance r is considered from. The empty space function considers some arbitrary point (i.e., not a point of the process $N(\Omega)$) located at some point x. Whereas both $G(r)$ and $K(r)$ refer to individual arbitrary points in the process, and are often referred to as Palm characteristics.

Point process statistics are typically defined with respect to $b(x, r)$, which refers to a sphere of radius r centred at some location x. For stationary processes, the functional statistics are given by,

$$\begin{aligned} H(r) &= 1 - \mathbf{P}(N(b(x,r)) = 0) \quad \text{for } r \geq 0 \\ G(r) &= \mathbf{P}(N(b(x,r)\backslash x) > 0) \quad \text{for } r \geq 0 \\ K(r) &= \tfrac{1}{\lambda}\mathbf{E}(N(b(x,r)\backslash x)) \quad \text{for } r \geq 0 \end{aligned}$$

and in the case of stationarity (that $N(\cdot)$ is invariant under translation) where x can be taken to be the origin, o,

$$\begin{aligned} H(r) &= 1 - \mathbf{P}(N(b(o,r)) = 0) \quad \text{for } r \geq 0 \\ G(r) &= \mathbf{P}(N(b(o,r)\backslash o) > 0) \quad \text{for } r \geq 0 \\ K(r) &= \tfrac{1}{\lambda}\mathbf{E}(N(b(o,r)\backslash o)) \quad \text{for } r \geq 0 \end{aligned}$$

Hence $H(r)$, the empty space function, is the complement of the void probability, that there exist no points within a distance r from x (or the origin, o). The nearest neighbour distribution function, $G(r)$, describes the distribution of distance from the points of $N(\cdot)$ to their nearest neighbour. It is the probability that $b(x, r)$ is not empty (not counting, $\backslash$, x itself). Ripley's K-function, $K(r)$, describes the expected number of points of $N(\cdot)$ within a distance r of an arbitrarily selected point of $N(\cdot)$ (not counting, $\backslash$, the point itself). This as, under the assumption of stationarity the sphere $b(x, r)$ can be centred at the origin (i.e., $b(o, r)$).

Estimation methodology for such point pattern characteristics is a topic beyond the scope of this thesis. However, one thing to note is their sensitivity to edge effects. Edge effects relate to the fact that $N(\cdot)$ is only observed within some bounded window, $\mathcal{W}$, yet $N(\cdot)$ is implicitly assumed to be operating outwith $\mathcal{W}$. This leads to biased estimators for the summary characteristics of $N(\cdot)$. For example, consider the nearest neighbour, y, of some observed point, x, of the process. Although $x \in N(\mathcal{W})$, this does not imply that $y \in N(\mathcal{W})$. Thus, if no information outwith $\mathcal{W}$ is available the true nearest neighbour of x may not be determined. This leads to biased estimators of summary characteristics unless some correction method is employed. In addition, the choice of observation window can affect inference. For instance if the window is too large, then one is characterising empty space near the

boundary. Broadly speaking there are three approaches which may be taken to deal with edge effects. These approaches include: (i) using buffer zones, (ii) making adjustments to account for unobserved events, and (iii) creating a torus of the considered domain. Respectively, these approaches: (i) create a buffer zone that includes all points within some distance d from the edge of the domain, and through conditioning on these points carries out all statistical analysis on those remaining; (ii) makes an intelligent "on average" adjustment to the estimate of interest; (iii) wraps a (rectangular) domain by identifying opposite edges essentially eliminating edges. Full details of these approaches are not discussed here, more comprehensive details can be found in Diggle (2013); Illian et al. (2008).

As mentioned above, comparison between the estimated functional characteristics $\hat{G}(r)$, $\hat{H}(r)$, and $\hat{K}(r)$, and their theoretical Poisson counterparts $G(r), H(r)$, and $K(r)$, typically enable some basic structure of the observed point pattern to be inferred. Consider Figure 1.2, which shows three realisations of point processes and their estimated Ripley's K-functions: a homogeneous Poisson process, a cluster process, and a regular process, respectively. Looking at the bottom plot of Figure 1.2 Ripley's K-function of the simulated patterns can be compared to infer their structure. The grey solid line in Figure 1.2 (bottom plot) shows Ripley's K function of a theoretical pattern of CSR. The estimated K-function of the simulated homogeneous Poisson process, shown in the top right plot of Figure 1.2, is shown by the solid black line. In addition, the dashed line in Figure 1.2 (bottom plot) is the estimated K-function for the simulated cluster process (top centre plot of Figure 1.2). The K-function for the regular process (top left plot of Figure 1.2), is shown by the dotted line.

Ripley's K-function for a homogeneous Poisson process, shown by both the solid grey line in Figure 1.2, represents the expected number of points within a distance r divided by the intensity, from some arbitrarily chosen point. If more points than expected, assuming a homogeneous Poisson process, were observed within these distances, then the pattern might be considered to be a realisation of a cluster process. Therefore, for a cluster process one would observe higher values of Ripley's K-function at the same distances. This is illustrated by the dashed line in Figure 1.2, which corresponds to the simulated cluster process (top centre plot of Figure 1.2). The empty space function $H(r)$ may be thought of as an analogue to the nearest neighbour distribution function $G(r)$ as they both describe distances from points to points in the point process. Yet, $H(r)$ considers the distance from some point not in the process, whereas $G(r)$ uses a typical point in the process as its reference point.

In addition to these aforementioned functional characteristics the literature commonly refers to the J-function, $J(r)$, and the pair correlation function (pair c.f), $g(r)$. The former is simply a function of $G(r)$, and $H(r)$:

$$J(r) = \tfrac{1-G(r)}{1-H(r)} \quad \text{for } r \geq 0, \quad \text{with } H(r) < 1.$$

This function will be ≥ 1 for regular patterns, and ≤ 1 for cluster processes. This due to

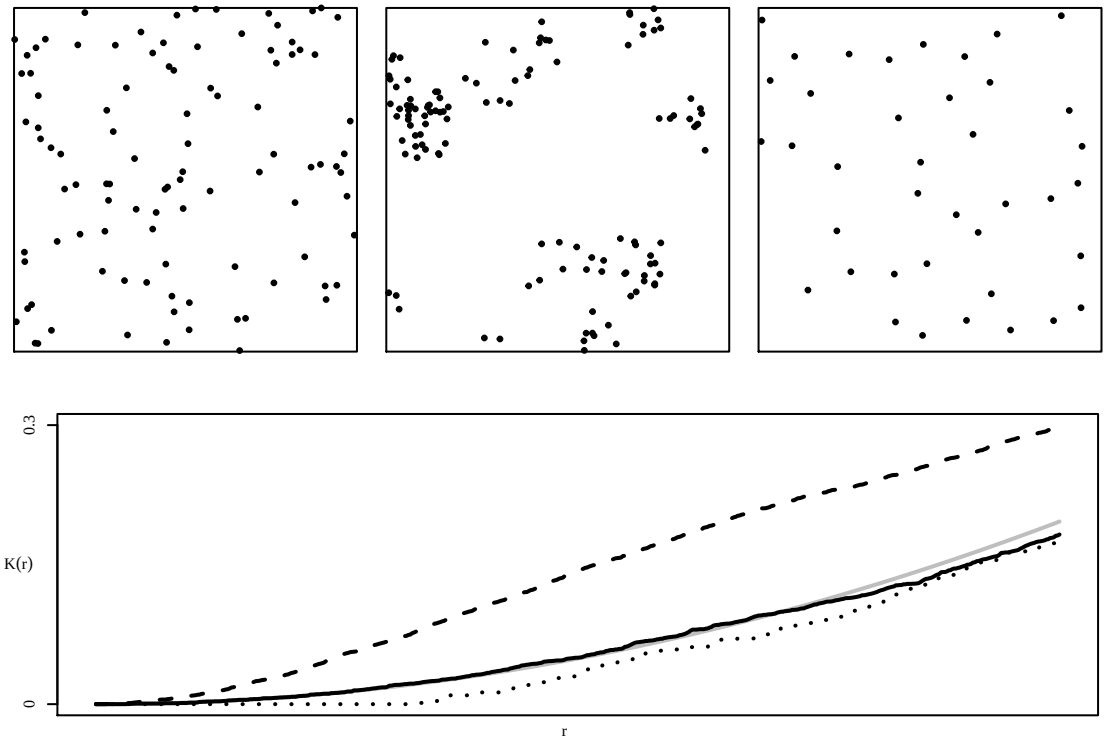


Figure 1.2: Three realisations of point processes, each with $\lambda = 100$, and their estimated Ripley's K-functions (bottom plot). From left to right respectively (top row): a homogeneous Poisson process, a cluster process, and a regular process. Each has been simulated to have the same expected number of points within the unit square. Comparing Ripley's K-function of theoretical CSR (grey solid line) to those for the simulated homogeneous Poisson process (solid black line), cluster process (dashed line), and regular process (dotted line) it is obvious that the cluster process has more points than those expected of a homogeneous Poisson process, and the regular process slightly fewer.

the subtle differences between the nearest neighbour function and the empty space function. For a regular process the inter-point distances are larger than the distances between typical points and other process points, therefore $G(r) \leq H(r)$ for $r \geq 0$, hence, $J(r) \geq 1$. For a cluster process the distances represented by $G(r)$ are short (due to the points being in clusters), whereas distances from the origin to the nearest cluster may be large. Therefore $G(r) \geq H(r)$ for $r \geq 0$ which makes $J(r) \leq 1$.

The pair c.f, $g(r)$, describes how the point density varies as a function of distance r from some arbitrarily chosen point. The pair c.f, $g(r)$, (in 2-dimensions) is related to the derivative of Ripley's K-function (with respect to r) by,

$$g(r) = \frac{K'(r)}{2\,\pi\,r}.$$

The function $g(r) = \frac{2\,\pi\,r}{2\,\pi\,r} = 1$ in the case of CSR. This is due to $K(r) = \pi\,r^2$ for a Poisson process (Illian et al., 2008, p. 77)). This reflects that the location of any point is independent of the locations of any other point. In general, $g(r)$ is proportional to the expected intensity of a point process at a distance r from an arbitrarily chosen point (under the assumption of stationarity). Specifically, $g(r) = \lambda_p(r)/\lambda$, where λ refers to the global point intensity. This

expected density with parameters $\boldsymbol{\theta}$, $\lambda_p(r;\boldsymbol{\theta})$, is known as the Palm i.f.

It is beyond the scope of this thesis to discuss more summary characteristics than those detailed above, but it suffices to say that they do exist, see Illian et al. (2008) for further details. However, given the characteristics of a point pattern typically the aim is to construct an appropriate model for the observed point pattern. Specifically, assuming that the data are a realisation of some point process. Section 1.2.2 introduces such models, and sketches the computational burden which is typical to fitting spatial models.

1.2.2 Point process models

Using functional summaries of the data (e.g., those detailed above) may be useful in inferring the structure of the observed point pattern. Estimated functional characteristics of the data do inform as to what class of model might be better to explain the variation. Figure 1.2 illustrates the two main structural arrangements typically observed in point patterns, which deviate from CSR: clustering and regularity. Other structural arrangements do exist, one of which is discussed in a later chapter of this thesis, but are not well covered in the point process literature.

Point processes are mathematical models which describe the geometrical structure of point patterns. This section gives an overview of the most commonly considered point process models in point process literature. This forms a basis for the models and methodologies developed throughout the following chapters of this thesis.

A list (and by no means an exhaustive one) of other types of point process models is given below. Some of these types are discussed further in this thesis; others are purely mentioned for reference purposes only. For a more comprehensive list of point process models, see Illian et al. (2008).

- **Typical cluster processes:** Listed below are perhaps the most common point processes that give rise to clustered point patterns. All those listed here are considered and extended in this thesis.
  - **Neyman Scott point process (NSPP):** a type of cluster process where clusters of points (the number per cluster being a Poisson random variable) are observed centred at some unknown (or hypothetical) locations governed by a Poisson point process. Classically, a Neyman-Scott point process (NSPP) is likened to an apple orchard, where only the fallen apples are observed (daughters), yet are sired by unobserved parents (apple trees). These parents are themselves realisations of a point process with homogeneous intensity. Two variants listed below are discussed further in Chapter 2 of this thesis:
    - **Modified Thomas process:** the points within each cluster are Gaussian distributed in space, centred at the parent's location (the pattern simulated

in the centre plot of Figure 1.2 is in fact a realisation of a Modified Thomas process)

* **Matérn cluster process:** the points within each cluster are uniformly distributed within a sphere of a defined radius, centred at the parent's location

– **Cox process:** here, $\lambda(s)$ is itself a realisation of some random process. This is sometimes termed a doubly stochastic process (Cox and Isham, 1980). The following listed process is considered in Chapters 4, 5, and 7 of this thesis.

* **log-Gaussian Cox process (LGCP):** a mathematically tractable type of Cox process, specifically it is a form of latent Gaussian model. A Cox process is a hierarchical process commonly used to model point patterns which exhibit irregular spatial structure. The assumed random process typically represents some unobserved environmental process. Given this intensity process, the observations, the points, are scattered independently (Rathbun, 1996; Møller et al., 1998). Assuming that clustered point pattern data are realisations of a log-Gaussian Cox process (LGCP) is a common approach in many fields, such as ecology (Illian et al., 2012a), epidemiology (Benes et al., 2002), environmental science (Møller and Díaz-Avalos, 2010), and international relations (Zammit-Mangion et al., 2012).
This framework assumes that the latent process follows a Gaussian distribution. More particularly that the intensity measure, $\Lambda(\mathbf{s})$, is given by,

$$\Lambda(\mathbf{s}) = \exp(\mathbf{z}(\mathbf{s})) = \int_{\Omega} \lambda(\mathbf{s}) d\mathbf{s}. \tag{1.2}$$

This transformation of the intensity function is both practical and functional. The stochastic structure, the Gaussian Random Field (GRF), approximates the intensity field of the Cox process, and, cannot be negative. A realisation of a LGCP has a Poisson number of points with mean $\Lambda(\mathbf{s})$.

- **Typical repulsion processes:** The other most commonly considered geometrical structure of a point pattern is regularity (i.e., repulsion). Although not directly considered in this thesis, it would be remiss of this introduction to not mention perhaps the most commonly discussed processes which give rise to regular point patterns[1]. Therefore, a brief summary of Matérn hardcore processes is given below.

– **Matérn hardcore processes:** these result from applying some thinning rule—some procedure leading to the deletion of points—to a homogeneous Poisson point process

[1]If the readers of this thesis are familiar with point process literature then these point processes are technically a type of Gibbs process, a class of point processes that model interaction among points.

- **Matérn I hardcore process:** this is obtained by deleting every point with its nearest neighbour closer than some given hard core distance (the pattern simulated in the right hand plot of Figure 1.2 is in fact a realisation of a Matérn I hardcore process)
- **Matérn II hardcore process:** this is obtained by deleting a point with its nearest neighbour closer than some given distance, if a randomly associated mark given to that point is larger than its neighbour

These two processes are of particular note, as some rule, or process, is deleting points. This is akin to some sort of hierarchical structure, whereby some process is governing the structure of the observed points. This is similar in some ways to a NSPP where this process is deleting points, not propagating them. Specifically, there are assumed to be unobserved parents which dictate the structure of the observed pattern, resulting in areas devoid of points as opposed to clusters of points. This parallel is mentioned specifically to aid the reader in the understanding of the point process developed in Chapter 3.

The following section discusses a class of latent model; specifically their relevance in capturing physical processes and dependence. The use of latent models for point pattern data is well documented (Moller and Waagepetersen, 2004; Illian et al., 2008; Diggle, 2013). Here the latent structures are used to both account for the spatial and or temporal dependence. This, enables one to infer the unconsidered processes proffering the observed structure. Moreover, when considering geostatistical data—Section 1.1.2—this latent process represents the spatially and or temporally continuous process observed at the sampled locations.

Latent Gaussian models

Latent models (i.e., models where a hierarchical modelling framework is supposed), assume some structured process to account for the inherent spatial, temporal, or spatio-temporal dependence (Cressie and Wikle, 2011; Gelfand et al., 2010; Diggle et al., 1998). Consider some response variable given by y_i observed at some locations s_i $(i = 1, \cdots, N)$ that are realisations from some distribution. The structured linear predictor η_i is given by

$$\eta_i = \beta_0 + \sum_{j=1}^{n} \beta_j x_{ji} + \sum_{k=1}^{m} z^k(s_i) + \epsilon_i, \tag{1.3}$$

for $i = 1, \cdots, N$ (the number of observations). In Equation 1.3 above each $z^k(s_i)$ is a k^{th} $(k = 1, \cdots m)$ structured process accounting for either spatial or temporal dependence at location s_i. In addition, it incorporates variation unexplained by the considered covariates. That is, $z^k(s_i)$ can be thought of as a process which embodies the inherent structure in the data due to both unidentified covariates and the spatial, temporal, or spatio-temporal

dependence. Each coefficient β_j $(j = 1, \cdots, n)$ represents the linear effect of each covariates x_{ji}, β_0 is an intercept term, and ϵ_i is an unstructured term.

Formulating the model in Equation 1.3 in a hierarchical structure sets the data to be conditionally independent given the latent process. In a Bayesian framework hyperparameters, that is parameters of the latent field, given by $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ are also considered. Hence, using the same notation as Martins et al. (2013) the model is formulated as follows,

Observed data: $\mathbf{y}|\mathbf{z}, \boldsymbol{\theta}_1 \sim \pi(\mathbf{y}|\mathbf{z}, \boldsymbol{\theta}_1) = \prod_{i=1}^{N} \pi(y_i|z_i, \boldsymbol{\theta}_1)$

Latent field: $\mathbf{z}|\boldsymbol{\theta}_2 \sim \pi(\mathbf{z}|\boldsymbol{\theta}_2) = N(\mathbf{z}; \boldsymbol{\mu}(\boldsymbol{\theta}_2), \mathbf{Q}^{-1}(\boldsymbol{\theta}_2))$

Hyperparameters: $\tilde{\boldsymbol{\theta}} \sim \pi(\tilde{\boldsymbol{\theta}})$

Above $N(\mathbf{z}; \boldsymbol{\mu}(\boldsymbol{\theta}_2), \mathbf{Q}^{-1}(\boldsymbol{\theta}_2))$ signifies a multivariate Gaussian distribution for the latent field given its parameters, with a vector of means $\boldsymbol{\mu}$ and precision matrix $\mathbf{Q}$. Each $z(s_i)$ in Equation (1.3) above is assumed to be a GRF. To calculate this random effect distribution, the multivariate log-Gaussian density given by,

$$-\frac{1}{2}\left(n\log(2\,\pi) + \log(|\mathbf{Q}^{-1}|) + [z(s) - \mu_f]^T\,\mathbf{Q}\,[z(s) - \mu_f]\right) \tag{1.4}$$

needs to be maximised. Note that due to the dense $n \times n$ covariance matrix $\boldsymbol{\Sigma} = \mathbf{Q}^{-1}$, maximising the log-likelihood given by Equation 1.4 is an operation of $\mathcal{O}(n^3)$, and is therefore very computationally expensive. Section 1.3.1 below discusses conditional independence properties of the latent GRF which leads to a Gaussian Markov Random Field (GMRF) due to the sparsity of the precision matrix $\mathbf{Q}$, reducing the order of operation to $\mathcal{O}(n^{3/2})$.

1.3 Fitting Latent Gaussian Models

In order to estimate parameters, $\boldsymbol{\theta}$, of the above mentioned point process models, one may consider a maximum likelihood based method, a least squares type estimation method such as minimum contrast, or a Bayesian approach. The first approach assumes that the model parameters are non-random. In particular, the log-likelihood function is maximised to yield parameter estimates, $\hat{\boldsymbol{\theta}}$, which best fit the data. An approximate likelihood technique, where the likelihood function is not known but approximated may also be used. This technique is considered for Cox processes, see Møller and Waagepetersen (2007), and in the case of NSPPs, see Tanaka et al. (2008). The latter case is based on the pair c.f (Section 1.2). Further details of this approximate likelihood method are given in Chapters 2, and 3, where the method is used to consider the spatial structure of histopathology data.

In some situations Bayesian approaches are more frequently taken to estimate parameters of point process models. In contrast to the maximum likelihood approach, a Bayesian approach considers the parameters to be random variables themselves following some probability distribution of their own. In particular, although the observations are still modelled

by some probability distribution, $L(\cdot)$, this is now considered to describe the conditional distribution of the observations given the parameters, $\boldsymbol{\theta}$. Moreover, the parameters $\boldsymbol{\theta}$ are themselves modelled by prior distributions, $\pi(\boldsymbol{\theta})$. Thus, the uncertainty relating to the parameters $\boldsymbol{\theta}$ is given by the posterior distribution, after using the evidence from the data, which is proportional to the product $\pi(\boldsymbol{\theta})\, L(\cdot)$. Here, the characteristics of the respective marginal distributions—obtained by integrating out over the posterior distribution with respect to all other parameters—enable the estimates $\hat{\boldsymbol{\theta}}$ to be inferred.

Typically, obtaining the marginal posterior distributions is burdensome, as generally this may not be available in closed form. Therefore other ways of gleaning information regarding the parameters are required. The most well known way of doing so is by Markov Chain Monté Carlo (MCMC), a simulation based methodology which samples from a probability distribution based on a Markov chain (Kendall et al., 2005; Lunn et al., 2012). In summary, this approach is extremely flexible with perhaps one drawback: a high computational burden. Although MCMC methods are extremely flexible, they are typically too computationally expensive to deal with complex dependencies in spatial point processes. As such, new methodology to enable the fitting of complex random effect models has arisen. One way to speed up computation is to utilise an automatic differentiation tool. This technique exploits the computational advantage which arises from numerically evaluating the derivative of the supplied function. One implementation of automatic differentiation is the Automatic Differentiation Model Builder (ADMB) project that implements reverse mode automatic differentiation to fit non-linear statistical models. In addition, this has been extended to incorporate random effects (ADMB-RE) (Fournier et al., 2012). Within the ADMB framework MCMC can easily be implemented. Due to the exploitation of the C++ language, implementing MCMC like this is much faster than the traditional Bayesian inference Using Gibbs Sampling (BUGS) language (Spiegelhalter et al., 1996; Lunn et al., 2012). Taking a similar approach to ADMB, the `R` package `TMB` (Template Model Builder) evaluates and maximises the Laplace approximation of the marginal likelihood in the fitting of complex nonlinear random effect models (Kristensen et al., 2016). Such methodology has seen use in the fitting of models to—but not exclusively to—animal movement data (Albertsen et al., 2015). Further details pertaining to the methods mentioned briefly above are beyond the scope of this thesis; the references provided will provide a far more comprehensive detailing of each approach mentioned. This thesis in many chapters utilises Laplace Approximations in order to fit complex spatial and spatio-temporal point process models. This, mainly through the use of the Integrated Nested Laplace Approximation (INLA) framework (see Section 1.3.1). However, a TMB based approach is also utilised in some instances. Further details of the INLA approximation technique are given below and utilised in Chapters 4, 5, and 6. In particular, the INLA approach is designed to fit latent Gaussian models making it particularly applicable to fitting LGCP models (Section 1.2.2).

1.3.1 Fitting Latent Gaussian Models using Integrated Nested Laplace Approximations

The INLA approach is designed to fit latent Gaussian models, discussed in Section 1.2.2 above, where the latent Gaussian fields have conditional independence properties. The conditional independence of the latent Gaussian field translates into a sparse precision matrix, which is important to the computational efficiency of the INLA approach. Letting $\mathbf{y}$ denote the observed data, and $\mathbf{z}$ denote the latent Gaussian field with parameters $\boldsymbol{\theta}$ the posterior distribution of the random field is given by,

$$\begin{aligned}\pi(\mathbf{z}, \boldsymbol{\theta}|\mathbf{y}) &\propto \pi(\boldsymbol{\theta})\pi(\mathbf{z}|\boldsymbol{\theta}) \prod_{i=1}^{N} \pi(y_i|z_i, \boldsymbol{\theta}) \\ &\propto \pi(\boldsymbol{\theta})|\mathbf{Q}(\boldsymbol{\theta})|^{\frac{N}{2}} \exp[-\frac{1}{2}\mathbf{z}^t\mathbf{Q}(\boldsymbol{\theta})\mathbf{z} + \sum_{i=1}^{N} \pi(y_i|z_i, \boldsymbol{\theta})],\end{aligned}$$

using the INLA approach a Gaussian approximation $\tilde{\pi}(\mathbf{z}, \boldsymbol{\theta}|\mathbf{y})$ to $\pi(\mathbf{z}|\boldsymbol{\theta}, \mathbf{y})$ is required. The marginals of interest required for this approximation are given by,

$$\begin{aligned}\pi(z_i|\mathbf{y}) &= \int \pi(z_i|\boldsymbol{\theta}, \mathbf{y})\, \pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \\ \pi(\boldsymbol{\theta}_j|\mathbf{y}) &= \int \pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}_{-j} \qquad j = 1, \cdots, m.\end{aligned}$$

Assuming that each data point $\{y_i, i = 1, \cdots, N\}$ depends on a single point in the field z_i, the INLA approach necessitates three main tasks; further details are given by Rue et al. (2009) and Martins et al. (2013):

1. The proposition of an approximation $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ to $\pi(\boldsymbol{\theta}|\mathbf{y})$
2. The proposition of an approximation $\tilde{\pi}(z_i|\boldsymbol{\theta}, \mathbf{y})$ to $\pi(z_i|\boldsymbol{\theta}, \mathbf{y})$
3. The exploration of $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ on a grid to integrate out $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_j$ in the approximations of $\pi(z_i|\mathbf{y})$ and $\pi(\boldsymbol{\theta}_j|\mathbf{y})$, respectively.

The approximation of $\pi(\boldsymbol{\theta}|\mathbf{y})$, where $\mathbf{x}^*(\theta)$ is the posterior mode, is

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \left.\frac{\pi(\mathbf{z}, \boldsymbol{\theta}, \mathbf{y})}{\pi_G(\mathbf{z}|\boldsymbol{\theta}, \mathbf{y})}\right|_{\mathbf{z}=\mathbf{x}^*(\theta)},$$

where, $\pi_G(\mathbf{z}|\boldsymbol{\theta}, \mathbf{y})$ is a Gaussian approximation. The INLA approach offers three different ways to approximate $\pi(z_i|\boldsymbol{\theta}, \mathbf{y})$:

1. Through using the marginals of the Gaussian approximation $\pi_G(\mathbf{z}|\boldsymbol{\theta}, \mathbf{y})$, computed when evaluating the approximation for the joint posterior of the hyper-parameters above.

2. Again perform a Laplace approximation. Thus, the approximation is given by,

$$\pi_{LA}(z_i|\boldsymbol{\theta},\mathbf{y}) \propto \left.\frac{\pi(\mathbf{z},\boldsymbol{\theta},\mathbf{y})}{\pi_{GG}(\mathbf{z}_i|z_i,\boldsymbol{\theta},\mathbf{y})}\right|_{\mathbf{z}_i=\mathbf{x}_i^*(z_i,\theta)},$$

   where $\pi_{GG}(\mathbf{z}_i|z_i,\boldsymbol{\theta},\mathbf{y})$ is the Gaussian approximation to $\mathbf{z}_i|z_i,\boldsymbol{\theta},\mathbf{y}$, $\mathbf{x}_i^*(z_i,\theta)$ is the modal configuration, and $-\mathbf{z}_i$ represents the vector $\mathbf{z}$ excluding its i^{th} element.

3. Performing a simplified Laplace approximation $\pi_{SLA}(z_i|\boldsymbol{\theta},\mathbf{y})$, which requires evaluating a Taylor expansion of the elements of $\pi_{LA}(z_i|\boldsymbol{\theta},\mathbf{y})$ up to third order.

Refer to Rue et al. (2009), Rue and Martino (2007), and Martins et al. (2013) for a detailed description of the procedures required for these approximations.

The use of latent models, Gaussian or otherwise, typically applies to situations where a process is defined on a continuous domain. Specifically, there is assumed to be some process (spatial field) underlying the observed data. Thus, a latent model defines the joint behaviour for all locations in the domain, $\mathbf{s} \in \Omega \in \mathbb{R}^d$; for physical world phenomena these are typically for $d = 2$. Considering the three types of spatial data discussed in Section 1.1, the spatial field, $\mathbf{z}(\mathbf{s})$, in the absence of any fixed effects, can be thought of as (i) a process accounting for the inherent neighbourhood structure of the data in the case of lattice data, where discretisation of the data occurs when recording data: (ii) an approximation of the entire process defined on some continuous domain Ω when considering geostatistical data; and (iii) in the case of point pattern data an approximation of the spatially varying intensity of points within a region.

As noted in Section 1.3.1, the sparsity of the precision matrix of the GMRF leads to efficient computations; however, as approximations of mechanisms operating on continuous domains are typically required in this context, it is beneficial to model the joint behaviour of this process for all locations. The section below details an approach which approximates this assumed continuously indexed spatial structure.

1.3.2 A Stochastic Partial Differential Equation (SPDE) model for the latent field

The use of a stochastic partial differential equation (SPDE) as an approximation to a GMRF was introduced by Lindgren et al. (2011). Here, weighted sums of basis functions are used to approximate the spatial random functions arising from the solution to the SPDE. Thus, the continuous interpretation of space is preserved, whilst benefiting from the computational advantages arising from the discrete Markovian structures of GMRFs.

Lindgren et al. (2011) show that—under conditions detailed below—the stationary solution to the SPDE given by,

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}} \tau\, \mathbf{x}(\mathbf{s}) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^d, \quad \alpha = \nu + \tfrac{d}{2}, \quad \kappa > 0, \quad \nu > 0, \tag{1.5}$$

is a GRF. In this SPDE, $\Delta = \sum_{i=1}^{d} \frac{d^2}{ds_i^2}$ is the Laplacian, κ is the spatial scale parameter, α controls the smoothness (i.e., ν measures the degree of smoothness of the process), and τ controls the variance. Here, $\mathcal{W}(\mathbf{s})$, is a Gaussian spatial white noise process. The stationary solution to the SPDE on $\mathbb{R}^d$ is a random field with a Matérn covariance function,

$$\mathcal{C}(\mathbf{x}(\mathbf{0}), \mathbf{x}(\mathbf{s})) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\kappa \parallel s \parallel)^{\nu} K_{\nu}(\kappa \parallel s \parallel), \tag{1.6}$$

where, κ and ν are as given in Equation (1.5), $\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)\kappa^{2\nu}(4\pi)^{\frac{d}{2}}} > 0$ is the marginal variance and K_ν is the modified Bessel function of second order.

Integer values of α in Equation (1.5) leads to continuous domain Markovian fields (Rozanov, 1982). In all models discussed in this thesis which utilise the SPDE approach, $\alpha = 2$. Lindgren et al. (2011) demonstrate that this provides discrete basis representations of $\mathbf{x}(\mathbf{s})$, and arguably is a more natural choice for two dimensional space (Whittle, 1954). Hence, the models discussed herein are a subset of those discussed in Lindgren et al. (2011) are built on finite basis representation given by,

$$\mathbf{z}(\mathbf{s}) = \Sigma_{k=1}^{m} \phi_k(\mathbf{s})\, w_k, \tag{1.7}$$

where the weight vector $\mathbf{w} = \{w_1, \cdots, w_m\}$ is chosen such that $\mathbf{z}(\mathbf{s})$ approximates the solution to the Equation (1.5). Here, each $\phi_k(\cdot)$ is a deterministic basis function defined by a triangulation of the domain, with m nodes. Through projection of the SPDE onto the basis representation, piecewise polynomial basis functions are constructed, enabling a Markov structure to be obtained and preserved when conditioning on local observations. Figure 1.3 shows a realisation of some point process (the point pattern shown in Figure 1.1 (c)) and the Delauney triangulation constructed over the domain; each triangle is termed a finite element. Thus, this triangulation of the domain, Ω, is the structure on which the finite element representation with piecewise linear basis functions is defined. The square plotting characters in Figure 1.3 illustrate the location of the observations, and the asterisks represent the m nodes of the triangulation. The shaded pyramid represents the value of the spatial field at that location; that is, one $z(s_k)$ in Equation (1.7). This is given by the corresponding weight, w_k, and the values in the interior of the triangle are determined by linear interpolation (Lindgren et al., 2011). Following the notation of Lindgren (2012), who define the sparse matrices $\mathbf{C}$, $\mathbf{G}_1$, and $\mathbf{G}_2$, the precision matrix for the weights in Equation (1.7) is given by,

$$\mathbf{Q} = \tau^2(\kappa^4\mathbf{C} + 2\kappa^2\mathbf{G}_1 + \mathbf{G}_2). \tag{1.8}$$

Choosing $\mathbf{z} \sim N(\mathbf{0}, \mathbf{Q}^{-1})$ produces a continuously defined function, $\mathbf{z}(\mathbf{s})$, that approximates the solution, $\mathbf{x}(\mathbf{s})$, to the SPDE given in Equation (1.5).

It should be noted that the construction of the triangulation is an influential step in model fitting (Lindgren, 2012). The spatial resolution of the triangulation affects inference; the construction of the triangulation should reflect the correlation range of the assumed latent

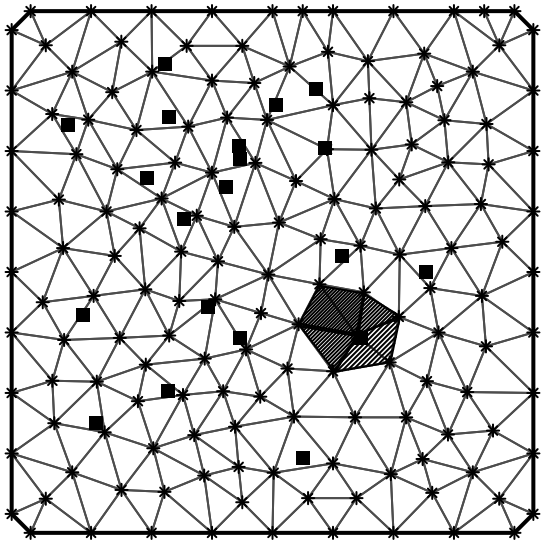


Figure 1.3: Triangulation of the domain for the point pattern shown in Figure 1.1 c) onto which the finite element methodoogy (FEM) representation of the continuous latent process is built. This triangulation discretises the domain, and facilitates the integration over the intensity surface. The plotting characters illustrate both the nodes of the triangulation, (asterisks), and the observations, (filled squares). The pyramid represents the finite element representation of the spatial field at that location, giving the logarithm of the intensity surface. The size of the triangles should represent the range of correlation in the data. Thus, if the spatial structure is assumed to be smooth the triangles can be larger (how large is discussed in detail later in the text). A projector matrix links the observations and the mesh, giving the neighbourhood structure of the observations. When modelling a point pattern the response vector is set to be 0 at the mesh nodes, and 1 at the observation locations. Thus, the SPDE model for the intensity surface essentially interpolates around these locations giving an estimate of the intensity surface over the domain.

process. If the spatial structure is smooth then the size of the triangles should reflect the little variation at neighbouring locations (i.e., can be comparatively larger). However, if there is a lot of local scale variation, the triangles must be small in order to capture this (i.e., comparatively small with respect to the size of the domain). In addition, it is customary that the triangulation (mesh) be built to extend beyond the region of interest in order to avoid boundary effects. However, in some instances, this area of increased uncertainty can be used to advantage. For example, consider analysis of marine creatures; in coastal areas their movements are more variable than in regions further offshore, and so utilising the boundary effect of the mesh on the coast can incorporate this increased variability. Sensitivity of inference to choice of triangulation is discussed later on in this thesis.

The remainder of this introduction outlines the methodologies detailed in each chapter of this thesis.

1.4 Thesis outline

This thesis focuses on fitting point process models using various methods. In particular, this thesis is aimed at extending classically considered classes of point process models. This, in order that further morphological structures, or dependencies, inherent in the considered point

patterns may be inferred. With the exception of Chapter 3—where the geometry of voids in a pattern is considered—all other chapters that consider point process models consider novel techniques in the fitting of cluster processes.

Following the acronym list at the end of this thesis, there is a list of uncommon mathematical notations used throughout this thesis, which the reader may be unfamiliar with. In addition, a species list, listing both the Latin and common names of any species data analysed follows the notation list.

Each chapter of this thesis has an accompanying `R` package for which the author of this thesis either extended the functionality of, or was sole author of. These packages are not yet available on CRAN. See the table below for information on where they can be downloaded, which chapters they are associated with, and the versions used in the writing of this thesis.

Package	**Functionality**	**Version**	**Download** (Linux)
`nspp`	Chapter 2	0.0.1	`https://github.com/cmjt/nspp`
`gapski`	Chapter 3	0.0.1	`https://github.com/cmjt/gapski`
`lgcpSPDE`	Chapters 4, 5, 6, and 7	0.0.1	`https://github.com/cmjt/lgcpSPDE`

The remainder of this section provides an overview of each chapter. It should be noted that in addition to chapters this thesis has been sectioned into parts, within which chapters are nested. This is to aggregate comparable methods; each part loosely follows on from the other by the incorporation of generalisations. These are aimed at developing novel techniques in the fitting of point process models that capture the complexities inherent in the motivating point patterns. Parts II to V form the core aspect of this thesis in the development of point process models. This introduction is nested within Part I. Part VI forms the discussion chapter.

1.4.1 Identifying unusual spatial structures

Part II chiefly focuses on using an approximate-likelihood approach in the fitting of point process models. Both Chapters 2 and 3 nested in this part approximate the likelihood function for the data through a technique utilising the Palm i.f. This enables estimation of the associated parameters.

The motivation for Part II is the analysis of the complex spatial structures in tumour tissue sections. These sections typically comprise both cancerous and noncancerous tissue structures, with regions of each intermixed in space (Mattfeldt and Fleischer, 2014). Thus, the locations of nuclei within the tissue can be considered as realisations of some point process. In each chapter the Palm i.f assumed to describe the intrapoint distances of the nuclei is derived facilitating maximisation of the approximate-likelihood function in each case.

When considering real world phenomena such as tissue morphometrics, identifying the

representative process is nigh on impossible. In particular, the morphology of cancerous tissue is a result of many complex processes describing the intermixing of cancerous and noncancerous cells. Thus, Chapters 2 and 3 fit both a cluster and a void process respectively, discussing their relevance in capturing the complexities inherent in the histopathology data considered.

The ideas presented in Chapters 2 and 3 have recently been submitted (Jones-Todd, Caie, Illian, Stevenson, Savage, Harrison, Bown, (2016)) to *The Annals of Applied Statistics*. The author of this thesis contributed to this work through deriving the Palm i.f for each of the processes discussed therein. In addition, the author extended functionality of the `R` package `nspp` (Stevenson, 2015), and developed the `R` package `gapski`. Both aforementioned packages provide software implementation for the approximate likelihood estimation of the processes discussed in Chapters 2 and 3 respectively.

An approximate-likelihood approach to estimate parameters of a Neyman-Scott point process and its application in predicting cancer patient survival

The cluster processes discussed in Chapter 2 constitute unobserved parent points, which themselves are realisations of a point process with homogeneous intensity. These parent points sire observable daughter points, which induces clustering. The intensity of the unobserved parent points can be thought of in terms of a latent process driving the arrangement of the clusters of daughters. This chapter defines Neyman Scott point processes as in Illian et al. (2008), and discusses two examples of the Neyman Scott point process—namely, the modified Thomas process, and the Matérn cluster process (see Section 1.2.2).

Chapter 2 takes an approximate-likelihood approach in fitting these cluster process to the histopathology data discussed above. This technique initially proposed by Tanaka et al. (2008) in the fitting of a modified Thomas process is based on the derivation of the Palm i.f. Moreover, functionality of the `R` package `nspp` (Stevenson, 2015) is extended by the author of this thesis to facilitate the simulation and fitting of a Matérn process model using the approximate-likelihood technique.

An approximate-likelihood approach to estimate parameters of a void point process and its application in predicting cancer patient survival

Little headway has been made into processes exhibiting void dynamics; similarities may be made between void processes, and aggregation processes (see Section 1.2.2). Processes giving rise to aggregated point patterns, such as hardcore Matérn processes, are a result of some dependent thinning operation (Matérn, 1986). Here points do not exist—due to the thinning operation—within a certain radius, r_0, of each point in the observed (thinned) point process. Despite the similarities at first glance, by definition hardcore processes describe regularity in point patterns due to some inhibitive interaction between all observed points (Stoyan,

1988). In contrast, realisations of void processes are a consequence of unobserved parents and therefore only exhibit areas of inhibition.

A void process assumes that unobserved parent points repel their observed daughters. Specifically, assuming a homogeneous pattern of points, parents expunge every daughter in a sphere of radius R centred at the parent's location. Hence, the remaining daughters are free from their parent's influence, and are thus realisations of a void process. A void process, therefore, is an example of a dependently thinned point process (Illian et al., 2008, p. 365).

Chapter 3 takes an approximate-likelihood approach in fitting a void process to the histopathology data discussed above. It is not claimed that derivation of a likelihood function pertaining to this process is impractical, as for the NSPP (Tanaka et al., 2008). However, this chapter demonstrates such an approach following on from Chapter 2 as a preliminary step in the fitting of a void process model derived herein. In addition, functionality of the `R` package `gapski`—developed by the author of this thesis—is introduced. This package simulates and fits void process models using the approximate-likelihood technique.

1.4.2 Accounting for spatial dependency in real world data—Modelling point pattern data using a log-Gaussian Cox process

Part III of this thesis has only one nested chapter (Chapter 4). This chapter focuses on the fitting of spatial point process models. Fitting such models is challenging, mainly due to the complex dependencies inherent in the data. Typically maximum likelihood techniques are only possible in the most straightforward of cases (e.g., the fitting of a Poisson point process model (Cressie and Wikle, 2011; Gelfand et al., 2010)). In some cases an approximate-likelihood method may be used in the fitting of Cox process models (Møller and Waagepetersen, 2007), or as Tanaka et al. (2008) and Stevenson et al. (*in submission*) do in the fitting of some classes of NSPP models. This technique is also employed in Chapters 2 and 3 of this thesis.

In recent years Bayesian techniques have facilitated the fitting of more complex spatial point process models. However, simulation based techniques such as MCMC, are typically computationally burdensome. Thus, methodology which is both flexible and computationally feasible is required. One such example is a technique based on Laplace approximations, INLA, is detailed by Rue et al. (2009) and summarised in Section 1.3.1.

Chapter 4 introduces the fitting of LGCPs models (see Section 1.2.2) given by Equation (1.2) using INLA. In addition, Chapter 4 notes the benefits of using a SPDE model, given by Equation (1.5) and detailed in Section 1.3.2, for the latent Gaussian field. The construction of a SPDE model for the latent field affords the spatial structure inherent in the process to be better captured. Moreover, the use of INLA circumvents the computational expense typically associated with such models.

The chief aim of Chapter 4 is in discussing the role of the latent field in a LGCP. Thus,

enabling the reader a clearer understanding of the extensions to this methodology derived in later chapters of this thesis. In addition, functionality of the `R` package `lgcpSPDE` developed by the author of this thesis is introduced in the simulation and fitting of LGCP models using the INLA–SPDE technique.

1.4.3 Shared stochastic structures in accounting for multiple dependencies

Part IV again utilises a INLA–SPDE approach (as introduced in Chapter 4) in the fitting of models to spatially indexed data. Both Chapters 5 and 6, nested in this part of the thesis, introduce a novel approach to account for multiple dependencies inherent in spatial data.

The motivation for Part IV is in the development of an extremely flexible class of latent Gaussian model which reflects the multiple dependencies inherent in many spatial data structures. In particular, Part IV proposes extending the methodology outlined in Chapter 4. This through taking an INLA based approach, detailed in Section 1.3.1, utilising a SPDE model. In addition, both Chapters 5 and 6 introduce some additional functionality of the `R` package `lgcpSPDE` for the simulation and fitting of marked LGCP models using the INLA–SPDE technique.

The employ of shared stochastic structures in a class of extremely flexible latent Gaussian models

Let $N(\cdot)$ be a point process, then typically any point $x \in N(\cdot)$ will have some characteristic (i.e., mark). Examples include the height of a tree, the age of an animal, or some classification of an event. This leads to a marked point process. Typically ignoring the information regarding the mark would affect inference. Considering a LGCP, then, some mark process—a process of which the marks are a realisation—may well be in part contributing to the intensity measure of the LGCP.

Chapter 5 introduces the use of shared stochastic structures whist dealing with point pattern data from the field of geology. Assumed latent fields are used to infer the processes which are inherent in multivariate data. In particular, in the context of a marked LGCP, they are assumed to capture inter mark dependency, conditional on point location.

The main focus of Chapter 5 is in incorporating both the dependencies between the marks and point locations, and the relationship between marks conditional on their locations in both a spatial and spatio-temporal context. In particular, the flexibility of the modelling procedure proposed in Chapter 5 is demonstrated through an applied example. Chapter 5 highlights that the application of such a class of model to seemingly very different real world marked point pattern data. This, enabling the mechanisms underlying the observed spatial structure of the data to be inferred.

The ideas presented in Chapter 5 are to be submitted (Jones-Todd, Illian, Marques (2017)) to the *Annals of Applied Statistics*. The author of this thesis contributed to this work by

developing the methodology detailed in that chapter. This, through formulating and the software implementation of the models discussed to the data detailed therein. The class of models described in Chapter 5 have the potential to unearth complex unconsidered physical phenomena in various scientific fields. For example, joint LGCPs are currently being used to predict wetland habitation of the *Grus grus* (crane) (Soranio-Redondo et al., sion), as well as in understanding the spatial dynamics of earthquake occurrence (Kirsty Bayliss, personal communication, November 2015).

The use of multiple latent Gaussian fields in inferring ecological interactions

In many contexts, there is a need to consider several variables simultaneously. This either because treating them in isolation would ignore the complexities inherent in the mechanisms of interest, or because the dependence structures operating amongst the variables are the very mechanisms of interest. For example, this is the case in the joint study of several diseases in epidemiology (Knorr-Held and Best, 2001; Kim et al., 2001; Langford et al., 1999) as well as in the context of considering relationships among multiple species sharing similar habitats. Since the species potentially compete for similar resources, modelling individual species ignoring any effects of this competition, and hence relevant biotic interactions, while only focusing on abiotic interactions would oversimplify the underlying dynamics (Boulangeat et al., 2012).

In complex ecosystems the dependence relationships are rarely simple (Dray et al., 2012). To provide a realistic modelling approach, and to enable the relevant dependence structures to be inferred, Chapter 6 proposes a multivariate approach. This approach not only accounts for dependencies in space but also over time. Simultaneously modelling several species employing a computationally efficient modelling approach (INLA–SPDE) allows one to capture both predator-prey and sympatric relationships.

The ideas presented Section 6.2 of Chapter 6 at the time of writing this thesis are under review after submission (Jones-Todd, Swallow, Illian, Toms (2016)) to the *Journal of the Royal Statistical Society (Series C)*. The author of this thesis contributed to this work by formulating and implementing the model discussed to the data detailed therein.

1.4.4 Second order structures of latent Gaussian random fields— Non-stationary log-Gaussian Cox processes

The final core part of this thesis, Part V, has only one nested chapter (Chapter 7). This chapter investigates what are herein termed second order structures of GRFs. In particular, through considering the form of the precision matrix of the GRF. The precision matrix (Equation (1.8)) of the SPDE model (Equation 1.5) assumes stationarity. This is consistent in the methods discussed in Chapters 2, 3, 4, 5, and 6 of this thesis.

Let $N(\cdot)$ be a point process, and $N_x(\cdot)$ be the resulting process from a shift of all points of $N(\cdot)$ by a vector $\mathbf{x}$. Then $N(\cdot)$ is stationary if

$$N(\cdot) \stackrel{\mathrm{d}}{=} N_x(\cdot).$$

That is if $N(\cdot)$, and $N_x(\cdot)$ have the same distribution, $\stackrel{\mathrm{d}}{=}$, for all translations $\mathbf{x}$.

Chapter 7 details methodology in order to fit LGCP models that do not assume stationarity. This is based on a reformulation of the precision matrix given by Equation (1.8), whilst maintaining the sparsity of $\mathbf{Q}$ (Equation (1.8)). In addition to the sparsity of this matrix aiding in circumventing the computational burden involved in fitting such models, the modelling procedure is further expedited through the use of Laplace approximations. This through using both the INLA and TMB approaches. The former approach is discussed in Section 1.3.1. The latter is based on an approach utilising the relatively recent developed software TMB, which employs Laplace approximations and Automatic Differentiation (AD). Further details are given in Chapter 7 including the required functionality of the `R` package `lgcpSPDE`—developed by the author of this thesis—in the simulation and fitting of such non-stationary LGCP models.

Part II

Identifying unusual spatial structures

Chapter 2

An approximate-likelihood approach to estimate parameters of a Neyman-Scott point process and its application in predicting cancer patient survival

2.1 Introduction

In their most common form, point pattern data (Illian et al., 2008; Gelfand et al., 2010) describe the location of physical events or objects. Such objects (or events) may exist at either a macroscopic or microscopic scale. Typically considered in two-dimensional space, these data often exhibit different structural properties. Such structure typically classified as either clustering, or inhibition (Moller and Waagepetersen, 2004; Illian et al., 2008) (see Section 1.2.2). A cluster point process, as the name suggests, results in a pattern where on a local scale the intensity of the process is high, due to grouping of points (e.g., cosmic clusters). It may be argued that clustering is a common structure exhibited by patterns formed by real world phenomena, or at least the most commonly noted. For example, the congregation of animals, the crowding of human inhabitation, the distribution of galaxies in space, typically exhibit clustering. Considering a much smaller scale, the morphology of cells within a tissue sample is herein assumed to be a realisation of a cluster process, namely a NSPP. These data comprise of both cancerous and non-cancerous cells, termed tumour and stromal cells respectively.

Consider a point pattern, a realisation of some point process, $N(\cdot)$, characterised by a vector of parameters $\boldsymbol{\theta}$. This vector may simply consists of the intensity, λ, as is the case of a

homogeneous Poisson process (see Section 1.2.2). Notwithstanding the appeal of this simple characterisation, rarely in the physical world are point patterns truly homogeneous. That is, point patterns exhibit a diverse range of configurations, which may operate at different scales. This chapter considers two types of cluster processes both of which have classically been computationally burdensome to fit. Now, the maximisation of the Poisson likelihood is relatively simple. This offers a basis for extension into parameter estimation for processes exhibiting more complex spatial structure. In particular through an approximate-likelihood method. This where a comparative likelihood—the Poisson likelihood with an appropriate intensity function—is used, which is thought to best approximate the process of interest.

Intuitively the inter-point distances in a point pattern hold information regarding the spatial arrangement of the points. Such, point process summary statistics (e.g., Ripley's K-, H-, and G-functions (Ripley, 1977)) are typically used to infer the spatial structure (see Section 1.2). Yet, such statistics, despite their fundamental importance to the characterisation of point process, in many instances fail to adequately capture the spatial structure of the point pattern. Thus, to adequately capture the spatial structure of the data, so that an appropriate intensity function can be utilised in the approximate-likelihood, a more befitting summary function is required. This function is herein assumed to be the Palm i.f; replacing the intensity of a homogeneous Poisson process with the Palm i.f generates the approximate-likelihood in this case.

Estimation methodology based on an approximate likelihood approach—for the parameters $\boldsymbol{\theta}$—initially proposed by Tanaka et al. (2008), and generalised by Stevenson et al. (*in submission*) is discussed in this chapter. This approach is then extended to enable estimation of the parameters of a Matérn cluster process (Section 2.3.1). This approximate-likelihood approach involves an extension of the traditional homogeneous Poisson likelihood, to incorporate information of the inter-point distances encapsulated by the Palm i.f.

The ideas presented and the Palm i.f derived in Section 2.3.1 has been submitted to the *Annals of Applied Statistics* Jones-Todd et al. (*in submission*). The author of this thesis has in addition worked on extending the `R` package `nspp`, forked from a package of the same name developed by Stevenson (2015), not yet available on CRAN but available at `https://github.com/cmjt/nspp`. The functionality of this package is given in Section 2.4. An application to using the functionality of the package is demonstrated in Section 2.5 which considered histopathology data pertaining to colorectal cancer patient data set.

2.2 Introducing the Neyman-Scott point process (NSPP)

NSPPs give rise to clustered point patterns. This Section defines NSPPs as Illian et al. (2008), and discusses two examples of the NSPP: the Matérn cluster process, and the modified Thomas process (see Section 1.2.2). Typically, the clustering induced by these processes can

be thought to describe the spatial distribution, say, of fallen apples around their (unobserved) parent tree. Specifically, the intensity of the unobserved parent points can be thought of in terms of a latent process driving the arrangement of the clusters of daughters. Yet, this hierarchical structure can be thought of in many contexts. For instance, under the name of trace-contrast models Stevenson et al. (*in submission*) assume that cetacean sightings from a two-plane aerial survey are a realisation of a NSPP. Then, through estimating the parameters of a NSPP cetacean density may be inferred. Specifically, the daughter points correspond to the detections of cetaceans observed by each plane, and the parent points are thought of as the average location of each cetacean over the course of the survey.

An NSPP is defined by three components: (i) the distribution of daughters sired by each (unobserved) parent, (ii) the distribution of the number of daughters sired by each (unobserved) parent, and (iii) the distribution of the distances between two independent daughter points. Parameters of a NSPP are given by $\boldsymbol{\theta} = (D, \phi, \gamma)$. Here, D is the intensity of the unobserved parent points, which are assumed to be realisations of a homogeneous Poisson process. The number of daughters sired by each parent is assumed to be IID from some discrete distribution characterised by the parameter ϕ, (e.g., in the case of a Poisson distribution ϕ would be the expectation). From here on, as only the modified Thomas and Matérn process are considered, which both are assumed to have a Poisson number of daughters, the parameter ϕ is given as ν (the expectation of a Poisson process). It should be noted however, that in the modelling approach taken by Stevenson et al. (*in submission*) there is no necessity that the number of daughters is Poisson (i.e., ϕ can be any parameter from some chosen distribution describing the distribution of the number of children). In addition, conditional on their parent's locations the daughters are scattered in space according to some distribution with parameter γ. In the case of a Matérn process $\gamma = R$, which corresponds to the radius of the sphere within which the daughters are scattered uniformly. In the case of the modified Thomas process, $\gamma = \sigma$ corresponds to the Gaussian dispersion (standard deviation) of daughters around their parents. Figure 2.1 shows realisations of both a Thomas[1]—left hand plot—and a Matérn[2]—right hand plot—process. Both patterns have been simulated, using functionality in the `R` package `nspp` with parameter vector $\boldsymbol{\theta} = (7, 8, 0.05)$. Figure 2.1 illustrates the differences between the Thomas and Matérn processes. This difference is due to the parameter pertaining to the dispersion of daughters around their parents. When $\gamma = R$ (the Matérn process) the daughters are much more tightly clustered around their parents, this as daughters can only be at most a distance R from their parents. However, in the case of

[1] running `set.seed(40)`
`sim.ns(pars=c("D"=7,"child.disp"=0.05,"child.par"=8),plot.points=TRUE)`, (functionality within the `nspp` package) will result in the simulated pattern shown

[2] running `set.seed(10)`
`sim.ns(pars=c("D"=7,"child.disp"=0.05,"child.par"=8),dispersion="uniform",plot.points=TRUE)`, (functionality within the `nspp` package) will result in the simulated pattern shown

the Thomas process the corresponding parameter value describes the Gaussian dispersion of daughters, and no such hard boundary constrains the daughters. As the patterns in Figure 2.1 are simulated, the parent locations are known—and are shown by the grey crosses. However, given daughter locations alone the data may be considered to be realisations of NSPPs with lower parent density, D, larger expected number of daughters, ϕ, and, larger dispersion γ. Here the data have been chosen to illustrate this problem as they are aggregated such that two parents in each case are close-by, hence attributing cluster identity to daughters is unclear. The approach taken here however does not attribute cluster identity to any daughter point. Only the distribution of distances between daughters is required. This through the Palm i.f, which leads to the estimation of $\boldsymbol{\theta}$ through an approximate-likelihood approach.

2.2.1 Parameter estimation for NSPPs

Due to the assumption of some unobserved structure of a NSPP (i.e., unobserved parent points), estimation of $\boldsymbol{\theta}$ presents complications (Tanaka et al., 2008). Illustrated by Figure 2.1, distinguishing between siblings and non-siblings is almost impossible, as obviously the locations and number of parents are unknown. To achieve this Tanaka et al. (2008) demonstrate that given the distribution of the distances between two independent daughter points an approximate-likelihood approach can be taken. Whereas Kopeckỳ and Mrkvička (2016) consider a minimum contrast and Bayesian estimation approach to achieve parameter estimation. This approach is based on the Palm i.f, a function of the parameters $\boldsymbol{\theta}$ (relating to each process), and distance, r (see Section 1.2). In summary, at distance r from a randomly selected point the Palm i.f, $\lambda_p(r, \boldsymbol{\theta})$, returns the expected density of the process. Such an approach has been previously considered by Tanaka et al. (2008), where the contrast process is constructed through taking the Euclidean differences between a selected point and all other points. These d-dimensional differences are assumed to be realisations of an inhomogeneous Poisson point process with intensity given by the Palm i.f discussed in Section 2.3. Thus, taking an approximate-likelihood approach, the estimator for $\boldsymbol{\theta}$ is given by,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \mathbf{r}),$$

which is evaluated through the numerical maximisation of $\log(L(\boldsymbol{\theta}; \mathbf{r}))$ with respect to $\boldsymbol{\theta}$. Here, $L(\boldsymbol{\theta}; \mathbf{r})$ is termed the Palm likelihood, which is discussed in Section 2.3.2.

In order to derive the Palm i.f one must consider the distribution of distances between two independent daughter points. Clearly this distribution of distances is dependent on the spatial scattering of daughter points around their parents, in the cases discussed here either normally or uniformly. The Palm i.f for a Thomas process is derived by Tanaka et al. (2008), however a similar approach as taken by Stevenson et al. (*in submission*) enables the derivation of the Palm i.f for a Matérn process. In each case the distance distributions can be partitioned into two aspects, those being the distribution of the distances between two independent daughter

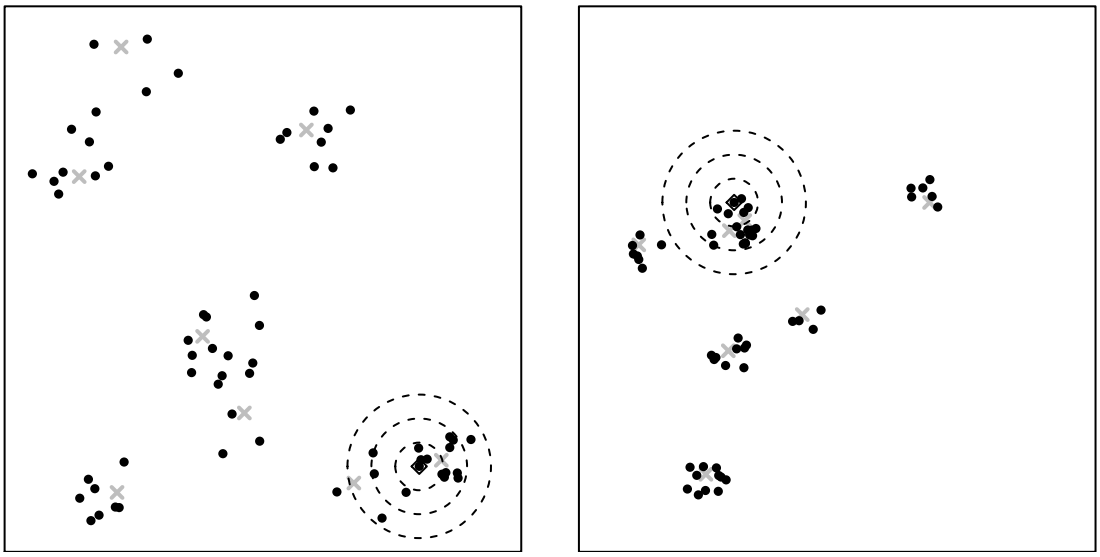


Figure 2.1: Two simulated NSPPs in the unit square; Thomas (left), and Matérn (right), where the unobserved parent points (grey crosses) "sire" the observed daughters (black dots). In each case parents are generated by a homogeneous Poisson process with intensity $D = 7$. The number of daughters sired by each parent are IID following a Poisson distribution with expectation $\phi = 8$. In each case $\gamma = 0.05$; thus, in the left hand plot daughters are dispersed around their parents due to a bivariate normal distribution, $\mathbf{N_2}(\mathbf{0}, \boldsymbol{\sigma^2 I_2})$, where $\gamma = \sigma = 0.05$, in the right hand plot daughters are uniformly distributed in a sphere with radius R around their parents, hence $\gamma = R = 0.05$.

points, sired by:(i) the same, and (ii) different parents. The spatial distribution of daughters sired by each parent enables the derivation of the distance distributions. By definition, in the case of the Matérn cluster process a Poisson number of daughters are uniformly distributed in a sphere around their parents, hence, as mentioned above, $\boldsymbol{\theta} = (D, \nu, R)$. Whereas the modified Thomas process has a Poisson number of daughters which are normally distributed around each parent, therefore, $\boldsymbol{\theta} = (D, \nu, \sigma)$. In both cases similar estimation procedures as those proposed by Tanaka et al. (2008), and Stevenson et al. (*in submission*) can be used, namely an approximate-likelihood approach.

2.3 The Palm intensity function

As discussed above the Palm i.f, $\lambda_p(r; \boldsymbol{\theta})$, is directly related to the pair c.f, $g(r; \boldsymbol{\theta})$, which describes how the point density varies as a function of distance r from an arbitrarily chosen point. Figure 2.1 illustrates how the point density in a sphere of radius r centred at an arbitrary observed point, changes as r increases. At short distances—small r (small circle in Figure 2.1)—the presence of a point suggests that there exist nearby points, hence $g(r; \boldsymbol{\theta}) \therefore \lambda_p(r; \boldsymbol{\theta})$ are high. The relationship between the Palm i.f, $\lambda_p(r; \boldsymbol{\theta})$, and the pair c.f, $g(r; \boldsymbol{\theta})$, is given by

$$\lambda_p(r; \boldsymbol{\theta}) = \tau\, g(r; \boldsymbol{\theta}),$$

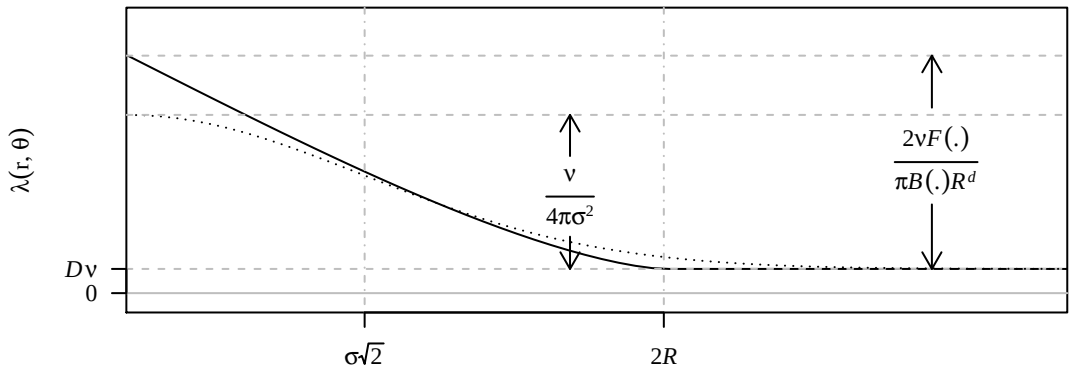


Figure 2.2: The functional form of the Palm i.f for both the Matérn (solid curve) and Thomas (dotted curve) variants of the NSPP in 2 dimensions. The horizontal asymptote is given by $D\,\nu$ for both processes. The difference between the horizontal asymptote and the $y-intercept$ is given by $\nu/(4\,\pi\,\sigma^2)$ for the Thomas process, and $2\,\nu\,F(\cdot)/(\pi\,B(\cdot)\,R^2)$ for the Matérn process. Here $F(\cdot) = {}_2F_1\left(\frac{1}{2}, \frac{1}{2} - \frac{\mathrm{d}}{2}, \frac{3}{2}, 1\right)$ denotes the hypergeometric function, and $B(\cdot) = B(\frac{3}{2}, \frac{1}{2})$ denotes the beta function. The point of inflection of the Gaussian term for the Thomas process is given by $\sigma\,\sqrt{(2)}$, for the Matérn process $2\,R$ is the point at which the Palm i.f decays to the horizontal asymptote.

which returns the expected density of a point process at a distance r from an arbitrarily chosen point. Here, τ is the global point density, the product of the parent density and the expected number of daughters.

In the case of CSR, $g(r, \boldsymbol{\theta})$ equals one, reflecting that the location of any point is independent of the locations of any other point. For a cluster process (e.g., a NSPP), $g(r, \boldsymbol{\theta}) \geq 1$. The pair c.f is high for small r, as at shorter distances from some arbitrarily chosen point we would expect a higher density of points. In addition, it should be noted that

$$\lim_{r\to\infty} g(r, \boldsymbol{\theta}) = 1 \quad \text{thus} \quad \lim_{r\to\infty} \lambda_p(r, \boldsymbol{\theta}) = \tau.$$

Specifically, the Palm i.f decays to the overall average intensity of the point process, τ, due to the behaviour of the pair c.f.

Under the conditions that (i) parent locations are realisations of a homogeneous Poisson process, (ii) the number of daughters sired by each parent are Poisson IID, and (iii) daughters are dispersed due to a bivariate normal distribution with a variance-covariance matrix of any (positive-definite) form, Tanaka et al. (2008) derive the analytical Palm i.f for a modified Thomas process, as,

$$\lambda(r; \boldsymbol{\theta}) = D\,\nu + \frac{\nu}{4\pi\sigma^2} \exp\left(\frac{-r^2}{4\sigma^2}\right), \tag{2.1}$$

where $D\,\nu = \tau$, the global point density. Thus, $\lambda(r; \boldsymbol{\theta})$ is the sum of the intensity of non-sibling points, $D\,\nu$, and a Gaussian function describing the intensity due to sibling points. Figure 2.2 shows the functional form of this Palm intensity (dotted line). This shows that as the distance r from a randomly selected point of the cluster process increases $\lambda_p(r, \boldsymbol{\theta})$ decays to the global point density $D\nu$ (the horizontal asymptote).

In comparison to the Thomas process the Palm i.f of the Matérn process (solid line in Figure 2.2) is initially a lot higher, and decays at a (much) faster rate (assuming the same

parameter values). This as, the parameter γ controls the dispersion of daughters around their parents (the radius of dispersion in a Matérn process and the spread in a Thomas process). Thus, the Matérn process, where $\gamma = R$ refers to a hard boundary, induces a (much) higher intensity at shorter distances. It must also be noted the similarities between the Palm i.f of the Matérn process and the void process (discussed in Chapter 3), namely the shape, is due to the geometry of the spatial structure.

Stevenson et al. (*in submission*) generalise the methodology proposed by Tanaka et al. (2008) for estimating $\boldsymbol{\theta}$ for an isotropic NSPPs based on a derivation of the Palm likelihood in the following ways:i) allowing NSPPs to be considered in any dimension, ii) allowing the numbers of daughters generated by a parent to follow any distribution, and iii) enabling the inclusion of information with regards to known siblings . Stevenson et al. (*in submission*) implement their described methods in a `R` package `nspp`[3] which is extended by the author of this thesis to incorporate estimation of $\boldsymbol{\theta}$ for the Matérn process, and used for the analysis of the data discussed in the following sections. The reminder of this section outlines the method, as presented by Stevenson et al. (*in submission*), and derives the extensions required to incorporate estimation of $\boldsymbol{\theta}$ for a Matérn cluster process, a class of NSPP.

2.3.1 Generalising the Palm intensity function

Noting that only the daughter points of the process are observed, the Palm i.f is derived through considering the expected number of daughters within a distance r of a randomly selected daughter. This is given by $\eta^d(r;\boldsymbol{\theta})$, and can be partitioned into points with the same, $\eta_s^{\mathrm{d}}(r;\boldsymbol{\theta})$, and different, $\eta_d^{\mathrm{d}}(r;\boldsymbol{\theta})$, parents. That is,

$$\eta^{\mathrm{d}}(r;\boldsymbol{\theta}) = \eta_d^{\mathrm{d}}(r;\boldsymbol{\theta}) + \eta_s^{\mathrm{d}}(r;\boldsymbol{\theta}).$$

Let $v^{\mathrm{d}}(r)$ and $s^{\mathrm{d}}(r)$ denote the volume and area of a d-dimensional hypersphere of radius r respectively, given by,

$$v^{\mathrm{d}}(r) = \frac{\pi^{\mathrm{d}/2}\, r^{\mathrm{d}}}{\Gamma(\frac{\mathrm{d}}{2}+1)}, \text{ and } \quad s^{\mathrm{d}}(r) = \frac{d}{dr} v^{\mathrm{d}}(r) = \frac{\mathrm{d}\,\pi^{\mathrm{d}/2}\, r^{\mathrm{d}-1}}{\Gamma(\frac{\mathrm{d}}{2}+1)}.$$

Then, following Stevenson et al. (*in submission*) the Palm i.f is given by,

$$\begin{aligned} \lambda(r;\boldsymbol{\theta}) &= \left(\frac{d}{dr}\eta^{\mathrm{d}}(r;\boldsymbol{\theta})\right)\frac{1}{s^{\mathrm{d}}(r)} \\ &= \left(\frac{d}{dr}\eta_d^{\mathrm{d}}(r;\boldsymbol{\theta}) + \frac{d}{dr}\eta_s^{\mathrm{d}}(r;\boldsymbol{\theta})\right)\frac{1}{s^{\mathrm{d}}(r)} \\ &= D\,E_c(\phi) + \frac{[E_s(\phi)-1]\, f_y^{\mathrm{d}}(r;\gamma)}{s^{\mathrm{d}}(r)}. \end{aligned} \tag{2.2}$$

Here $E_c(\phi)$ and $E_s(\phi)$ are the expected numbers of daughters sired by a randomly selected parent and the expected number of daughters sired by a parent of a randomly selected

[3]available from `https://github.com/cmjt/nspp`, all examples and applications shown in this chapter have been run using functionality of this version of the package.

daughter respectively, both characterised by the parameter ϕ. The subtleties of the differences are due to the homogeneity of the parent distribution (i.e., each parent is equally likely to be selected), however, a daughter from a larger cluster is more likely to be chosen. Letting Y denote the distance between two randomly selected siblings (i.e., daughters sired by the same parent), $f_y^{\mathrm{d}}(r;\gamma)$ is then the PDF of this random variable, which is characterised by the parameter γ pertaining to the dispersion of daughters around their parents.

The PDF of the random variable, Y, is derived by Stevenson et al. (*in submission*) in the case of the modified Thomas process, that is, where Y denotes the distance between two normally distributed siblings. In this case the PDF in Equation (2.2) is derived as,

$$f_y^{\mathrm{d}}(r;\gamma) = \frac{2^{1-\mathrm{d}/2}\, r^{\mathrm{d}-1} \exp(-r^2/(4\gamma^2))}{(\gamma\sqrt{2})^{\mathrm{d}}\, \Gamma(\mathrm{d}/2)}, \tag{2.3}$$

where, $\gamma = \sigma$, is the parameter describing the Gaussian dispersion of daughters around their parents. Upon substitution of Equation (2.3) into Equation (2.2), and noting that for the special case where the number of daughters sired by a randomly selected parent follows a Poisson distribution, $E_c(\phi) = E_s(\phi) - 1 = \nu$, the Palm i.f derived by Tanaka et al. (2008) (Equation (2.1)) is obtained.

Extending to the Matérn cluster process

Following the methodology outlined above, the only difference in the case of the Matérn process is the distribution of the distances between two randomly selected daughters sired by the same parent. Specifically, the daughters are now considered to be uniformly distributed around their parents. The parameter γ in Equation (2.2) now refers to the radius of the sphere, R, centred at a selected parent outwith which we do not observe sired daughters. Figure 2.1 illustrates the two types of NSPPs simulated with the same value of γ. Thus, for the same value of γ clearly the Palm i.f for the Matérn process is initially much higher, and decays at a much faster rate to the horizontal asymptote, $D\,\nu$. As illustrated in Figure 2.2, letting $\gamma = R$, $\lambda(r;\boldsymbol{\theta})$ is a continuous piecewise monotonic function of two sub-domains, $[0, 2R]$ and $[2R, \infty)$. The common endpoint of the sub-domains, $2R$, relates to the structure of the Matérn process, that is, the distance between two sibling daughters cannot be more than the diameter, $2R$, of a sphere centred at an unobserved parent. Thus, the probability of observing a sibling at a distance r from an arbitrarily chosen daughter pertains to the intersection of the hyperspheres centred at these points, $b(x, R) \cap b(y, R), x \neq y \in N$, where N is the cluster point pattern. Clearly when the distance between these points, $r \geq 2R$ then, the intersection ,$b(x, R) \cap b(y, R) = 0$.

The PDF, $f_y^{\mathrm{d}}(r;\gamma)$, of the distance between two siblings, assuming that the daughters are uniformly distributed in a d-dimensional hypersphere of radius R around their parents is derived by Tu et al., the hypergeometric function and integral representation are respectively given by,

$$
\begin{aligned}
f_y^{\mathrm{d}}(r;R) &= \frac{2\,\mathrm{d}}{B(\frac{\mathrm{d}}{2}+\frac{1}{2},\frac{1}{2})}\,\frac{r^{\mathrm{d}-1}}{R^{\mathrm{d}+1}}\left[{}_2F_1\left(\frac{1}{2},\frac{1}{2}-\frac{\mathrm{d}}{2},\frac{3}{2},1\right)R-{}_2F_1\left(\frac{1}{2},\frac{1}{2}-\frac{\mathrm{d}}{2},\frac{3}{2},\frac{r^2}{4\,R^2}\right)\frac{r}{2}\right],\\
&= \frac{2\,\mathrm{d}\,r^{\mathrm{d}-1}\int_{\frac{r}{2}}^{R}(R^2-x^2)^{\frac{\mathrm{d}-1}{2}}dx}{B(\frac{\mathrm{d}}{2}+\frac{1}{2},\frac{1}{2})\,R^{2\,\mathrm{d}}}.
\end{aligned}
\tag{2.4}
$$

Here $B(\cdot,\cdot)$ denotes the beta function, and ${}_2F_1(\cdot,\cdot,\cdot,\cdot)$ the hypergeometric function. Noting that for $\mathrm{d}=2$ and $\mathrm{d}=3$ the PDF in Equation (2.4) reduces to the forms given below which are equivalent to the PDFs of the distances between two randomly selected siblings given by in the respective dimensions.

1. for $\mathrm{d}=2$

   It should be noted that,

   $$
   \begin{aligned}
   \int_{\frac{r}{2}}^{R}(R^2-x^2)^{\frac{\mathrm{d}-1}{2}}dx &= \int_{\frac{r}{2}}^{R}(R^2-x^2)^{\frac{1}{2}}dx\\
   &= \frac{1}{8}\left(4\,R^2\sec^{-1}\left(\frac{2\,R}{r}\right)-r\,\sqrt{4\,R^2-r^2}\right).
   \end{aligned}
   $$

   therefore,

   $$
   \begin{aligned}
   \frac{2\,\mathrm{d}\,r^{\mathrm{d}-1}\int_{\frac{r}{2}}^{R}(R^2-x^2)^{\frac{\mathrm{d}-1}{2}}dx}{B(\frac{\mathrm{d}}{2}+\frac{1}{2},\frac{1}{2})\,R^{2\,\mathrm{d}}} &= \frac{4\,r\int_{\frac{r}{2}}^{R}(R^2-x^2)^{\frac{1}{2}}dx}{B(\frac{3}{2},\frac{1}{2})\,R^4}\\
   &= \frac{r\,\Gamma(2)}{\frac{1}{2}\sqrt{\pi}^2\,R^4\,2}\left(4\,R^2\sec^{-1}\left(\frac{2\,R}{r}\right)-r\,\sqrt{4\,R^2-r^2}\right)\\
   &= \frac{r}{\pi\,R^4}\left(4\,R^2\sec^{-1}\left(\frac{2\,R}{r}\right)-r\,\sqrt{4\,R^2-r^2}\right)\\
   &= \frac{r}{\pi\,R^4}\left(4\,R^2\cos^{-1}\left(\frac{r}{2\,R}\right)-r\,\sqrt{4\,R^2-r^2}\right),
   \end{aligned}
   $$

   noting that $\sec^{-1}(x)=\cos^{-1}\left(\frac{1}{x}\right)$, and as $B(\frac{3}{2},\frac{1}{2})=\frac{\Gamma(\frac{3}{2})\,\Gamma(\frac{1}{2})}{\Gamma(2)}$, recalling that $\Gamma(\frac{3}{2})=\frac{1}{2}\sqrt{\pi}$, $\Gamma(\frac{1}{2})=\sqrt{\pi}$, and $\Gamma(2)=1$.

2. for $\mathrm{d}=3$

   It should be noted that,

   $$
   \begin{aligned}
   \int_{\frac{r}{2}}^{R}(R^2-x^2)^{\frac{\mathrm{d}-1}{2}}dx &= \int_{\frac{r}{2}}^{R}(R^2-x^2)dx\\
   &= \frac{1}{24}\,(r-2\,R)^2\,(r+4\,R).
   \end{aligned}
   $$

therefore,

$$\begin{aligned}
\frac{2\,\mathrm{d}\,r^{\mathrm{d}-1}\,\int_{\frac{r}{2}}^{R}(R^2-x^2)^{\frac{\mathrm{d}-1}{2}}dx}{B(\frac{\mathrm{d}}{2}+\frac{1}{2},\frac{1}{2})\,R^{2\,\mathrm{d}}} &= \frac{6\,r^2\,\int_{\frac{r}{2}}^{R}(R^2-x^2)dx}{B(2,\frac{1}{2})\,R^6} \\
&= \frac{6\,r^2}{24\,B(2,\frac{1}{2})\,R^6}\,(r-2\,R)^2\,(r+4\,R)\,, \\
&= \frac{r^2\,\Gamma(\frac{5}{2})}{4\,R^6\,\Gamma(2)\,\Gamma(\frac{1}{2})}\,(r-2\,R)^2\,(r+4\,R) \\
&= \frac{r^2\,\frac{3}{4}\,\sqrt{\pi}}{4\,R^6\,\sqrt{\pi}}\,(r-2\,R)^2\,(r+4\,R) \\
&= \frac{3\,r^2}{16\,R^6}\,(r-2\,R)^2\,(r+4\,R) \\
&= \frac{3\,r^2}{16\,R^6}\left(R-\frac{r}{2}\right)^2\left(2\,R+\frac{r}{2}\right).
\end{aligned}$$

noting that $\Gamma(\frac{5}{2})=\frac{3}{4}\sqrt{\pi}$, $\Gamma(\frac{1}{2})=\sqrt{\pi}$, and $\Gamma(2)=1$.

Upon substitution of the PDF, given by Equation (2.4), into the Palm intensity function, given by Equation , as in the case of the modified Thomas process above, simplifications occur which circumvent numerical instability in $\lambda(r;\boldsymbol{\theta})$ at $r=0$. This as both the numerator and denominator in the second term contain the term r^{d-1}. Thus,

$$\begin{aligned}
\lambda(r;\boldsymbol{\theta}) &= D\,E_c(\phi)+\frac{[E_s(\phi)-1]\,f_y^{\mathrm{d}}(r;\gamma)}{s^{\mathrm{d}}(r)}, \\
&= D\,E_c(\phi)+\frac{[E_s(\phi)-1]\,2\,\mathrm{d}}{B(\frac{\mathrm{d}}{2}+\frac{1}{2},\frac{1}{2})}\,\frac{r^{\mathrm{d}-1}}{R^{\mathrm{d}+1}}\,\frac{\Gamma(\frac{\mathrm{d}}{2}+1)}{\mathrm{d}\,\pi^{d/2}\,r^{\mathrm{d}-1}} \\
&\times\left[{}_2F_1\left(\frac{1}{2},\frac{1}{2}-\frac{\mathrm{d}}{2},\frac{3}{2},1\right)\,R-{}_2F_1\left(\frac{1}{2},\frac{1}{2}-\frac{\mathrm{d}}{2},\frac{3}{2},\frac{r^2}{4\,R^2}\right)\,\frac{r}{2}\right], \\
&= D\,E_c(\phi)+\frac{2\,[E_s(\phi)-1]}{B(\frac{\mathrm{d}}{2}+\frac{1}{2},\frac{1}{2})R^{\mathrm{d}+1}}\,\frac{\Gamma(\frac{\mathrm{d}}{2}+1)}{\pi^{d/2}} \\
&\times\left[{}_2F_1\left(\frac{1}{2},\frac{1}{2}-\frac{\mathrm{d}}{2},\frac{3}{2},1\right)\,R-{}_2F_1\left(\frac{1}{2},\frac{1}{2}-\frac{\mathrm{d}}{2},\frac{3}{2},\frac{r^2}{4\,R^2}\right)\,\frac{r}{2}\right], \qquad (2.5) \\
&= D\,E_c(\phi)+\frac{2\,[E_s(\phi)-1]}{R^{\mathrm{d}+1}\,\pi^{\frac{\mathrm{d}}{2}}}\,\frac{\Gamma(\frac{\mathrm{d}}{2}+1)^2}{\Gamma(\frac{\mathrm{d}}{2}+\frac{1}{2})\,\Gamma(\frac{1}{2})} \\
&\times\left[{}_2F_1\left(\frac{1}{2},\frac{1}{2}-\frac{\mathrm{d}}{2},\frac{3}{2},1\right)\,R-{}_2F_1\left(\frac{1}{2},\frac{1}{2}-\frac{\mathrm{d}}{2},\frac{3}{2},\frac{r^2}{4\,R^2}\right)\,\frac{r}{2}\right], \\
&= D\,E_c(\phi)+\frac{2\,[E_s(\phi)-1]}{R^{\mathrm{d}+1}\,\pi^{\frac{1}{2}(\mathrm{d}+1)}}\,\frac{\Gamma(\frac{\mathrm{d}}{2}+1)^2}{\Gamma(\frac{\mathrm{d}}{2}+\frac{1}{2})} \\
&\times\left[{}_2F_1\left(\frac{1}{2},\frac{1}{2}-\frac{\mathrm{d}}{2},\frac{3}{2},1\right)\,R-{}_2F_1\left(\frac{1}{2},\frac{1}{2}-\frac{\mathrm{d}}{2},\frac{3}{2},\frac{r^2}{4\,R^2}\right)\,\frac{r}{2}\right].
\end{aligned}$$

noting that $B(x,y)=\frac{\Gamma(x)\,\Gamma(y)}{\Gamma(x+y)}$, and $\Gamma(\frac{1}{2})=\sqrt{\pi}$.

2.3.2 The Palm likelihood

In order to estimate $\boldsymbol{\theta}$, the likelihood is derived as

$$L(\boldsymbol{\theta};\mathbf{r}) = \left(\prod n\,\lambda(r;\boldsymbol{\theta})\mathrm{d}r\right)\; exp\left(-n\int_0^t \lambda(r;\boldsymbol{\theta})\,s^{\mathrm{d}}(r)\mathrm{d}r\right), \tag{2.6}$$

where the integral is a volume of d-dimensions around the intensity axis, simplified as

$$\begin{aligned}\int_0^t \lambda(r;\boldsymbol{\theta})\,s^{\mathrm{d}}(r)\mathrm{d}r &= \int_0^t \left(D\,E_c(\phi) + \frac{[E_s(\phi)-1]\,f_y^{\mathrm{d}}(r;\gamma)}{s^{\mathrm{d}}(r)}\right) s^{\mathrm{d}}(r)\\ &= D\,E_c(\phi)\,v^{\mathrm{d}}(t) + [E_s(\phi)-1]\,F_y^{\mathrm{d}}(t;R),\end{aligned}$$

where $F_y^{\mathrm{d}}(t;R)$ is the CDF of the distance between two randomly uniformly distributed siblings, $f_y^{\mathrm{d}}(r;R)$. Thus, this Palm likelihood is a generalisation of the special cases derived by Tanaka et al. (2008) and Stevenson et al. (*in submission*).

This CDF given by Tu and Fischbach (2002), $F_y^{\mathrm{d}}(t;R)$, is given by;

$$F_y^{\mathrm{d}}(t;R) = \frac{t^{\mathrm{d}}}{R^{\mathrm{d}}}\left(1 - \frac{B_\alpha\left(\frac{1}{2},\frac{\mathrm{d}}{2}+\frac{1}{2}\right)}{B\left(\frac{1}{2},\frac{\mathrm{d}}{2}+\frac{1}{2}\right)}\right) + 2^{\mathrm{d}}\,\frac{B_\alpha\left(\frac{\mathrm{d}}{2}+\frac{1}{2},\frac{\mathrm{d}}{2}+\frac{1}{2}\right)}{B\left(\frac{1}{2},\frac{\mathrm{d}}{2}+\frac{1}{2}\right)}, \tag{2.7}$$

where $B_\alpha(\cdot,\cdot)$ and $B(\cdot,\cdot)$ are the incomplete beta function and the beta function respectively.

2.4 Extending the `nspp` package

This section focuses on the implementation of an approximate-likelihood method in order to estimate parameters of a Matérn process. Specifically through the maximisation of the (approximate) likelihood given by Equation (2.7). This is achieved by extending the `R` package `nspp` written by Stevenson (2015), which implements the general estimation method for the modified Thomas process described in the previous sections. It should be noted that most of the functionality given below is based on an early version of that package. The reminder of this section introduces, and describes extensions to the `nspp` package in the simulation and fitting of Matérn point processes. In addition, examples for various functions found within the extended package are provided.[4]

2.4.1 Simulating a Matérn cluster process–the `sim.ns()` function

The `sim.ns()` function has been extended to simulate a Matérn cluster process in d-dimensions, an example of a simulated Matérn process using `sim.ns()` in 2D is provided with the package–found in the exported object `example.mat.2D`.

The function has been extended from the version given by Stevenson et al. (*in submission*) through the inclusion of an additional argument allowing the user to choose the type of NSPP

[4]The extended functionality of the `R` package is available at `https://github.com/cmjt/nspp`.

they wish to simulate. As a result of this the argument pertaining to the vector of named parameters also is changed. The additional arguments are

- `dispersion`: a character either `gaussian`, the default, or `uniform` which specifies the type of NSPP the user wishes to simulate (i.e., a modified Thomas or Matérn cluster process respectively).
- `pars`: a vector of model parameters, $\boldsymbol{\theta}$. Must contain named elements `D`, `child.disp`, and `child.par`–corresponding to D, γ, and ϕ, respectively. The named element `child.disp`[5] pertains to the parameter σ when `dispersion = 'gaussian'`, and the parameter R when `dispersion = 'uniform'`.

The other main arguments are,

- `lims`: a matrix with d rows giving the limits of a dimension of the observation window $\mathcal{W}$. NB: due to the periodic boundary constraints only a (hyper)rectangular $\mathcal{W}$ is permitted.
- `rchild`: a function that simulates random variables from the distribution describing the number of daughters sired by their parents, with first argument the number of random variables to generate, and the second the parameter ϕ.

Additional plotting functionality can be called using arguments of the function, for example using `plot.points=TRUE` will plot the simulated points, and `plot.empirical=TRUE` will plot the empirical palm function Stevenson et al. (*in submission*).

Simulating data with $D = 7$, $R = 0.05$ and the number of daughters sired by each parent following a Poisson distribution with mean, $\phi = 8$ in 2 dimensions is accomplished through running the following code:

```
points<- sim.ns(pars=c(D=7, lims = rbind(c(0, 1), c(0, 1)),
                child.disp=0.05,child.par=8),dispersion="uniform",
                rchild=rpois)
```

The object `points`[6] is then a matrix with n rows and 2 columns, where each row pertains to the locations of the simulated daughters.

2.4.2 Fitting a Matérn cluster process–the `fit.ns()` function

Estimation of $\boldsymbol{\theta}$ for a Matérn cluster process is achieved via the `fit.ns()`. Maximising the log of the Palm likelihood, given in Equation (2.6), numerically through use of the `optim()` function using the `'L-BFGS-B'` method.

[5] Stevenson et al. (*in submission*) previously had this parameter as `sigma`.

[6] Running `set.seed(10)` prior to executing this code will result in the same simulated pattern as shown in the right hand plot of Figure 2.1 and the same `points` object used throughout this section.

The function has two equivalent arguments to the `sim.ns()` function, namely `dispersion`[7] and `lims`. The other two main arguments are,

- `R`: the truncation distance, a distance within which a daughter is thought to be from its parent,
- `child.dist`: a list containing information about the distribution of the number of daughters sired by each parent. The named components of this list must be[8]:
  - `mean`: a function that takes ϕ as its only argument and returns the mean, $E_c(\phi)$.
  - `var`: a function that takes ϕ as its only argument and returns the variance, $V_c(\phi)$.
  - `sv`: the start value of ϕ for the optimisation algorithm.
  - `bounds`: a vector containing bounds for the parameter ϕ.

Thus, to fit a Matérn process to the simulated point pattern, `points` above one might run the following code in `R`.

```
child.dist = list(mean = function(x) x, var = function(x) x,
                                  sv = 8, bounds = c(1e-6, nrow(points)))

fit.mat.pois.2D <- fit.ns(points = points, lims = rbind(c(0, 1), c(0, 1)),
                            R = 0.5, dispersion="uniform",
                            child.dist = child.dist)
```

The object `fit.mat.pois.2D` is a list containing components which provide information about the fitted model, accessing this information is simple through the use of existing `R` utility functions. For example, the fitted Palm i.f—dashed line—in Figure 2.3 can be plotted through calling `plot()` on the returned fitted model object. However, in order to illustrate the model fit here (Figure 2.3) the fitted Palm i.f—dashed line—is overlaid onto the empirical—solid line—Palm i.f (see Stevenson et al. (*in submission*) and (Fewster et al., 2016) for details of the empirical Palm i.f). As mentioned in Section 2.3 above the Palm i.f decays to a horizontal asymptote, the global point density $D\nu$ of the pattern.

[7] In order to fit a Thomas process, this argument should be changed to ``gaussian'', the default.

[8] It should be noted that in order that a Matérn process is simulated the number of daughters sired by a parent must follow a Poisson distribution, thus the parameter ϕ below equates to the parameter ν discussed above. However, the functionality discussed here does allow for the user to specify any distribution pertaining to the numbers of daughters sired by each parent, yet only one parameter, ϕ can be estimated. This extension is a result of the work done by Stevenson et al. (*in submission*) for his `R` package of the same name, which is extended herein to include estimation and simulation of a Matérn cluster process.

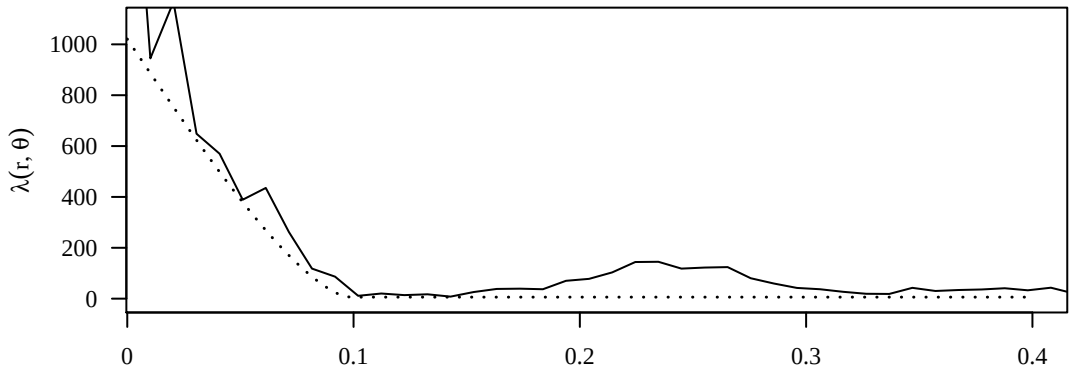


Figure 2.3: Both the empirical—solid lines—and fitted—dashed lines—Palm intensities for the simulated pattern, `points`, above. It is evident that this function is a piece-wise continuous function with two sub-domains $(0, 2R], [2R, \text{inf})$. As can be seen in each case the Palm intensities decay to the global point density of the process.

2.4.3 Variance estimation–the `boot.ns()` function

Following the methodology of Stevenson et al. (*in submission*) variance estimation is carried out using a parametric bootstrap. This is straightforward utilising the function `boot.ns()` implemented by Stevenson et al. (*in submission*), due to the extensions having already been made to the `sim.ns()` and `fit.ns()` functions. The main arguments of `boot.ns` are simply the fitted model, supplied as the first argument, `rchild`, equivalent to the argument of `sim.ns()` with the same name, and `N`, the number of bootstrap resamples. Thus, a bootstrap can be carried out through running the following `R` code,

```
boot.mat.fit <- boot.ns(fit.mat.pois.2D, rchild = rpois, N = 1000)
```

A call to `summary()` will now print out the estimated parameters $\boldsymbol{\theta}$ and their associated standard errors.

```
summary(boot.mat.fit)
## Coefficients:
##            Estimate Std. Error
## D          8.369725     5.0420
## child.disp 0.044518     0.0306
## child.par  8.231053     5.2283
```

2.5 Application to histopathology data

This section focuses on the application of the above developed methodology. An approximate-likelihood approach is taken in order to estimate parameters of a NSPP, of which the locations of cell nuclei are assumed to be a realisation. The histopathology data considered pertains to

42 patients drawn from a wider dataset of a pan-Scotland cohort of patients with colorectal cancer (see Caie et al. (2014) for a full description of the data). The data set is divided into two groups: 23 patients that died from colorectal cancer, and 19 that did not die from colorectal cancer. Each patient had up to 15 histopathology tissue sections processed, which were analysed automatically by an imaging algorithm (described in Caie and Harrison (2016)). In summary, distinct regions in the tissue section were first divided into four types: (i) tumour, (ii) stroma, (iii) necrosis/lumen and (iv) no tissue . The positions of all cell nuclei were automatically identified (see Caie and Harrison (2016)) and exported by the image analysis software Definiens, see Figure 2.4. The centres of each nucleus provide us with our point patterns of tumour and stroma cells. The positions of all cell nuclei were identified as detailed in Caie and Harrison (2016).

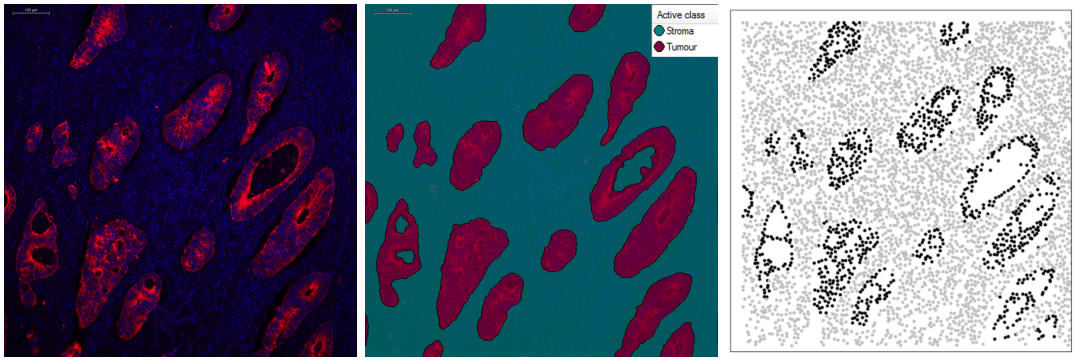


Figure 2.4: Illustration of the image analysis of one patient's slide which enables the pinpointing of nuclei. Left: Composite immunofluorescence digital image showing Tumour (red), Stroma (green) and all nuclei (blue). Middle: Image analysis mask overlay from automatic machine learnt segmentation of the digital image. Tumour (purple), stroma (turquoise), necrosis (yellow). Right: Point pattern formed by the nuclei of the tumour (black) and stroma (grey) cells shown in the previous two images.

Typically, pathologists assess tissue sections such that the tumour may be classified as relating to a particular stage—early, or late etc.—of cancer progression. Associating these stages to patient outcome faces many challenges, including subjectivity in grading (Wöhlke et al., 2011), as well as the multi-scale nature of the tissue structures being assessed. Therefore, if only in order to purely avoid subjectivity an automated classification approach offers many benefits.

The spatial arrangement of cells—both cancerous and non-cancerous (stromal)—within histopathology tissue sections are unusual in their distribution, due to the intermixing of the two different cell types. Contained in the spatial structure, is information, which may act as an indicator of patient survival. Thus, a method which enables the complexities inherent in the data to be captured motivates its application to other histopathology data. Considering the spatial morphology of nuclei within a tumour requires consideration to be given to both cell types. The use of point process statistics to consider the morphology of cancerous tissue is not new; Mattfeldt and Fleischer (2014) considered point process intensity and functional

summary statistics (i.e., Ripley's K-function and $g(\cdot)$) for two patient groups, those whose cancer had and had not metastasised (a major indicator of patient survival). Yet, no differences between patient groups with respect to either spatial statistics were found. The authors themselves conceptualised that the statistics used were not satisfactory in summarising the unusual spatial structures arising from the spatial intermixing of tumour and stroma.

This chapter assumes that the locations of cell nuclei identified on the histopathological slides detailed above are realisations of a Thomas process. The aim being that the parameters of the process may act as an indicator of patient survival. Through assuming that the spatial distribution of the above mentioned nuclei are realisations of a NSPP, the concept of parents takes a rather abstract slant. Typically, parents are assumed to be some unobserved physical phenomenon Wiegand et al. (2007). In the context of histopathology however, a NSPP can be thought of as a process which somehow generates the morphology of the tissue.

This methodology assumes that(i) parent points are realisations of a homogeneous Poisson point process with intensity, D, (ii) the number of daughters sired by each parent follow a Poisson distribution with mean ν; and (iii) the Gaussian dispersion of daughters around each parent has a variance-covariance matrix of the form $\Sigma = \sigma^2\mathbf{I}_2$, where $\mathbf{I}_2$ is a 2×2 identity matrix. Thence, the parameters to be estimated are given by $\boldsymbol{\theta} = (D, \nu, \sigma)$—parameters of a modified Thomas process.

Due to the structure of the data comprising two membership groups of tumour and stroma, and in order to estimate $\boldsymbol{\theta}$, information with regards to the heterogeneous nature of the tissue structure is required. This information can inform the starting value for the fitting procedure to facilitate convergence. Using a technique derived from graph theory this may be thought of as a class cover problem (Cannon et al., 1998), whereby an algorithmic approach is taken in order to form two disjoint groups. Generally, proximity graphs offer invaluable information as to the neighbourhood structure, and interaction among two or more classes. Consider Figure 2.5, which shows two classes of points—diamonds and circles—in layman's terms the aim of one particular proximity graph, a Class Cover Catch Digraphs (CCCD) (for one set of point, the diamonds in this case) is to find the minimum number of circles which encompass all the same class of points without encircling a point of the opposing class. However, in this context restricting the pattern such that—what may be thought of as—cluster density of one class of point is forced by the locations of the second class makes an unnecessary assumption about interactions between classes. Thus, rather than using characteristics of the CCCD (see the following section) to classify the patterns formed by nuclei location, they are used to inform the likely maximum distance one would expect a daughter (nuclei) to be from its (abstract) parent. This in order to aid optimization of the model fitting procedure. Note, assuming that each class of point is a realisation of a Thomas process does not suppose for instance that the spatial region occupied by a cluster of tumour cells cannot also be occupied by stroma cells.

Information pertaining to the intensity of the unobserved parent points can be gleaned

without making judgements with regards to the ancestry of the daughters. Therefore, herein information obtained from a CCCD for each slide (the average radius of the circles Figure 2.5) is used to inform plausible distances beyond which observed points cannot be siblings, as well as plausible parent densities (the number of the circles in the CCCD). The fitting procedure is carried out using the function `fit.ns()` (described above) from the `R` package `nspp`, with the argument `dispersion = ‘‘gaussian’’`. This in order that a Thomas process is fitted and not a Matérn process. A Thomas process is assumed as it is perhaps a more sensible choice. This due to it allowing more flexibility in the spatial dispersion of cells around their abstract parents. The fitting is carried out using plausible starting values (as required when using a maximum likelihood approach) obtained from the constructed CCCD. The following section gives a more formal description of CCCDs, after which it is investigated whether the spatial distribution of cells can aid in discriminating between patient survival outcomes.

Class cover catch digraphs

A simple graph is a pair of sets (V, E), where E is the set of edges connecting the set of $V = v_i, ..., v_n$ vertices. A directed graph (digraph), is a graph where the edges are ordered and as such are denoted by (V, D). Data random digraphs have often been used in the context of spatial pattern recognition (Marchette, 2005), a subtype of these graphs named CCCDs (Priebe et al., 2001) are used to classify data in multiple dimensions (DeVinney and Priebe, 2006). In general proximity graphs offer invaluable information as to the neighbourhood structure, and interaction among two or more classes, however the choice of graph can often bias the spatial analysis (Rajala, 2010). Hence, constructing a CCCD for each slide informs the heterogeneous structure of the tissue. Specifically, the distance beyond which two daughters cannot be sisters.

Consider the situation where there are observations from two classes $X = x_1, ..., x_n, Y = y_1, ..., y_m \in \mathbf{R}^q$, the class cover problem aims to find the smallest collection of spheres centred at observations in X such that every observation in X is in at least one of the spheres and no observation in Y is in any sphere. If one now considers a collection of sets $B_1, B_2, ...$ with associated base points $t_1, t_2, ...$ a catch digraph (V, D) has a directed edge from v_i to v_j if and only if $t_j \in B_i$. Therefore for any sets $X, Y \in \mathbf{R}^q$ the CCCD to be the catch digraph formed by base points $x_i \in X$ and the associates sets $B_i = z \in \mathbf{R}^q : d(x_i, z) < d(x_i, Y)$.

One may naïvely suppose that the sets $B_1, B_2, ...$ of a CCCD and the parent density, D, relating to one class of point proffer the same information. However, here it is conceptualised that these parent points are in fact abstract concepts pertaining to the process of cell division and interaction within the micro-environment of the tumour tissue. Thus, it is not assumed that the parents of each class, which would be the base points $x_i \in X$ and $y_i \in Y$ in the CCCD methodology, are observed. Thence, a robust measure which represents the latent process proffering the spatial distribution of points, taking into account the interactions between the

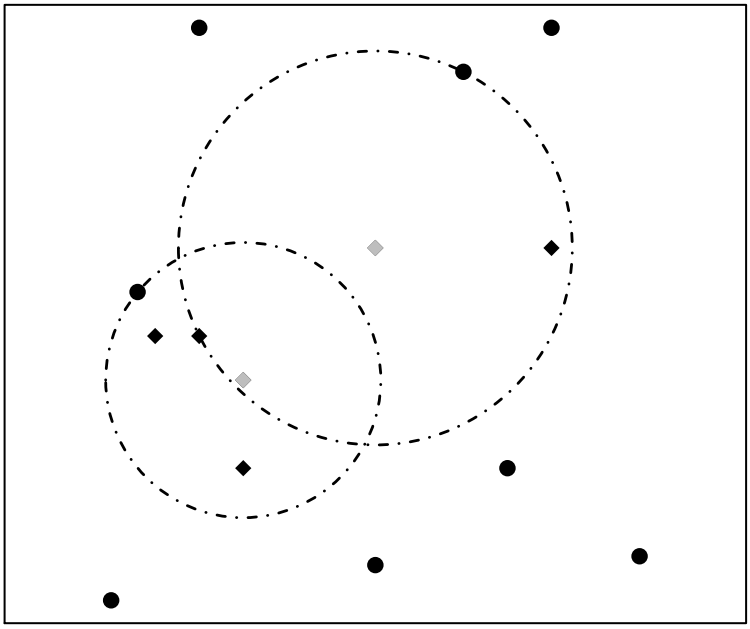


Figure 2.5: An example of a CCCD. The two types of plotting characters indicate two different sets of points. Let the set X be the diamonds, and the set Y be the filled circles. The base points $x_1, x_2 \in X$ are shown by the grey diamonds. Consider the spheres centred at these base points, every observation in X is contained in at least one of these spheres, and no observation of Y is. A directed edge from these base points forms a CCCD.

classes is required, hence, the CCCD–NSPP approach is employed.

2.5.1 Results

It is herein assumed that the patterns formed by the tumour and stroma nuclei respectively are realisations of Thomas processes. Functionality—the `fit.ns()` `R` function described above—within the `R` package `nspp`[9], is used to estimate $\boldsymbol{\theta}$ pertaining to each pattern. It should be noted that by default the argument `dispersion = ‘‘gaussian’’`, thus to fit a Thomas process this argument does not need to be specified. In addition, in order to fit the models it is required that we indicate the maximum distance we would expect a daughter to be from its parent, hence a reflection of the intra-cellular structure. This information is not supplied heuristically but inferred from the use of CCCD, detailed above. The estimation of $\boldsymbol{\theta} = (D, \phi, \gamma)$ leads to the analytic derivation of daughter intensity, herein referred to as δ.

In order not to make any distributional assumptions, a hierarchical bootstrap—using 1000 simulations—is required due to there being multiple slides per patient. Specifically, individual patterns are not assumed to be independent, and the bootstrapped samples are devised to have the same dependence structure as the original sample. Table 2.1 contains the medians and bootstrapped quantiles of the distributions of both the parent, $\hat{D}_{\cdot}$, and daughter, $\hat{\delta}_{\cdot}$,

[9]This package has been extended to incorporate parameter estimation for the Matérn process, and is available from `https://github.com/cmjt/nspp`

		Quantile of Order	
	Median	2.5%	97.5%
$\hat{D}_t^a$	91.6681	59.1620	262.9865
$\hat{D}_t^d$	65.0291	47.2256	89.9160
$\hat{D}_s^a$	89.4620	41.1063	154.9419
$\hat{D}_s^d$	21.4108	8.7270	42.9449
$\hat{\delta}_t^a$	2064.0908	1576.2919	2564.8912
$\hat{\delta}_t^d$	1233.6765	966.8923	1550.2797
$\hat{\delta}_s^a$	2964.7871	1685.0153	4095.1930
$\hat{\delta}_s^d$	1044.6352	740.6736	1515.6027

Table 2.1: The medians and hierarchical bootstrapped quantiles (based on 1000 simulations) of the distributions of both the parent, $\hat{D}_\cdot^\cdot$, and daughter, $\hat{\delta}_\cdot^\cdot$, intensities for each of the point patterns formed by tumour and stroma nuclei—subscript t and s respectively—of either patients who lived or died—superscript a and d respectively, assuming a Thomas cluster process.

densities, pertaining to each tumour and stroma point pattern. Here the subscripts t and s refer to the tumour and stroma nuclei respectively. The superscripts d and a refer to whether the patients died or survived respectively.

The results in Table 2.1 show that: (i) the values of both parent and daughter intensities are higher for patients who lived than for patients who died in both stroma and tumour data, (ii) the differences in the distribution of both parent and daughter densities between patients who lived and patients who died are larger in the stroma data than in the tumour data, (iii) for tumour data, there is no overlap in the distributions between patients who lived and patients who died in the daughter densities, while parent density distributions overlap, (iv) for stroma data, there is no overlap in the distributions between patients who lived and patients who died in daughter densities. As mentioned above, the meaning of a parent in this setting is rather abstract. The concept of daughters rather more obvious. However, one may postulate that the parameters of the assumed NSPP describe the spatial structure of the cells. The validity of assuming that the cell locations are a realisation of a NSPP is tested and discussed below.

In order to assess the predictive power of each of these classifiers (estimates from each point process) receiver operating characteristic (ROC) curves are used. Specifically, the means of each classifier were calculated per patient for use in a logistic regression (in predicting patient outcome). Figure 2.6 shows the ROC curves for each classifier along with the associated area under the ROC curve (AUROC) and leave one out cross validation (CV) scores. An AUROC score is a measure of classifier performance. To put the calculated AUROC scores in context, it should be noted that a score of 0.5 would indicate that the classifier does no better than random at distinguishing between patients who died or survived. In addition, a AUROC

score of 1 would indicate perfect separation between each patient group (i.e., no overlap in distribution of our classifiers). The CV scores are an estimate of the test error for the logistic regression models using each classifier. From Figure 2.6, notably, daughter density of the assumed Thomas process for the stroma patterns is considered the best predictor, with the highest AUROC score (0.71) and lowest comparative CV score (0.22). This indicating that stroma patterning is more informative than tumour patterning in the context of predicting patient survival.

This section chiefly presented a novel application of spatial statistics to capture the spatial arrangement of cells using a NSPP, illustrating the value of analytical methods to understand the spatial structure of cancer. It is the predictive ability of the classifiers mentioned above that is of particular interest. This demonstrates that from the spatial structure of cancerous tissue one can quantitatively infer a patient's outcome. It should be noted that the interpretations of the classifiers are based on the the suitability of the point processes in adequately describing the structure of the cell nuclei. It was assumed that the cell nuclei were a realisation of a NSPP. By comparing the spatial patterning of the observed data and date simulated from the fitted model, the suitability of the model can be informally assessed. Here, the empty space function was used for this comparison. Recall from Section 1.2.1 that the empty space function, $H(r)$ is the complement of the void probability ($\mathbf{P}(N(b(x,r)) = 0)$ for $r \geq 0$). This, therefore, in 2D, describes the probability that the disc $b(x,r)$ is not empty. In the stationary case, as is assumed here, the disc can be centered at the origin (i.e., $b(o,r)$).

Figure 2.7 shows the empty space functions for the modified Thomas process for the slide shown in Figure 2.4. Each solid line represents the empty space function for the fitted model; the dotted lines are the empty space functions estimated from patterns simulated with the respective estimated parameters. From Figure 2.7 we can see that a NSPP is unlikely to be the true process that describes the spatial patterning of cells. Figure 2.7 indicates that pattern formed by tumour cells is slightly less likely to have nearby points at the same distances as a NSPP. A comparison of the empty space function for both the tumour and stroma cell patterns to that of the homogeneous Poisson process (thick dashed lines) reveals that at short distances the cells exhibit some regularity, whereas only at longer distances does the pattern exhibit clustering. This is expected as in both cases the points represent cell nuclei, and due to the size of the cells at short distances no other cells (points) can exist. It should also be noted that the stroma cell pattern more quickly approaches that of a homogeneous Poisson process. From Figure 2.7 this is the case at a distance of 3% of the width and height of the slide ($r = 0.03$). This seems reasonable from the pattern of stroma cells shown in Figure 2.4.

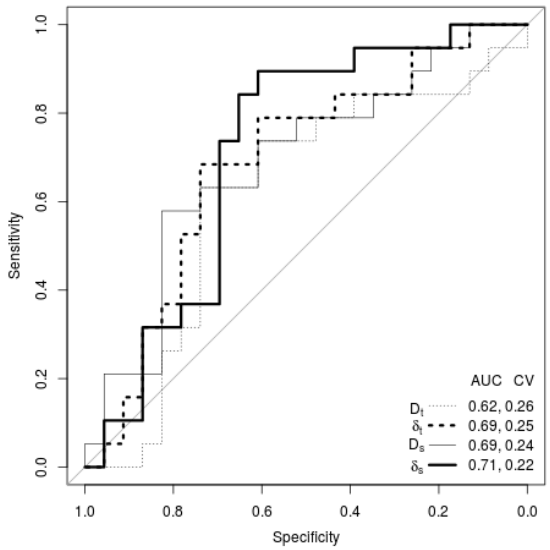


Figure 2.6: Plot of the ROC curves for each of the measures per patient. Legend includes the AUC (area under the curve) score for each curve and the CV (leave one out cross validation) score for each classifier (i.e., $D.$, and $\delta.$ parent, daughter densities of a Thomas process for tumour and stroma patterns—subscripts t and s respectively). The ROC curve plots false positive rates (specificity), against true positive rate (sensitivity); therefore, a diagonal line of 45° (grey solid line in the plot) would indicate that the classifier were equally likely to predict the outcome as not.

2.6 Discussion

This chapter has detailed an extension to the methodology proposed by Tanaka et al. (2008) (who estimates parameters of a Thomas process), in order that parameters of a Matérn process may be estimated. This extension is solely down to the derivation of the distribution of distances between two arbitrary chosen daughter points, which are now assumed to be uniformly distributed—as opposed to normally distributed as in the case of the Thomas process—around their parents.

The approach proposed by Stevenson et al. (*in submission*) shows that the method of Tanaka et al. (2008) leads to a strong observed bias for a Thomas process. This is effectively eradicated through considering the expected number of siblings—$E_s(\phi)$—rather than the expected number of daughters—$E_c(\phi)$—in the second term of the Palm i.f, as is illustrated in Equation (2.2). Thus, through taking a similar approach only the PDF of the distance between two normally distributed siblings is required in order to extend such methodology to estimate parameters of the Matérn process ($\boldsymbol{\theta} = (D, \phi, R)$, as in Equation (2.5)). The derivation of this PDF is detailed in this chapter. Moreover, functionality of the `R` package `nspp` in simulating and fitting this sub-class of NSPP is illustrated.

The remainder of this discussion considers the estimation procedure for the Matérn process described above, and the application of Thomas process models to histopathology data (pertaining to patients with colorectal cancer), which could potentially improve the understanding of the spatial structure of cells, aiding in predicting patient survival.

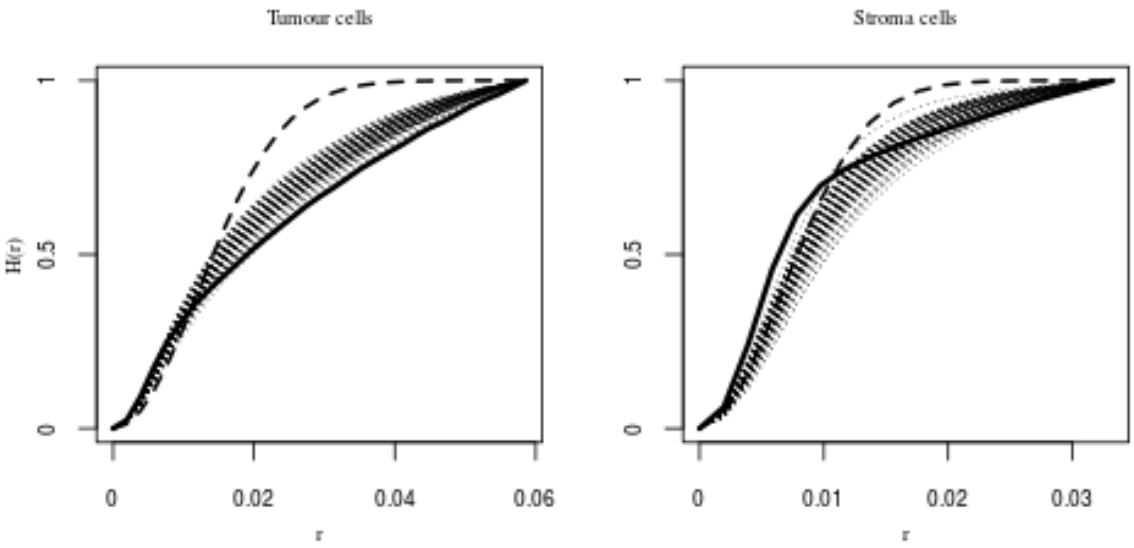


Figure 2.7: Figure showing for one patient's slide (Figure 2.4) the fitted empty space function, $H(r)$, by the solid line, and those for 100 simulated patterns (simulated using the estimated parameter values for each process), dotted lines. The thick dashed line is the theoretical empty space function for a homogeneous Poisson process. The distance r represent the percentage of the slide's width and height scaled to be $\in [0, 1]$.

2.6.1 Estimation for the Matérn process

Perhaps one of the most widely used `R` packages for point process statistics is `spatstat` (Baddeley and Turner, 2005); functionality within this package allows the fitting of a Matérn cluster process—a sub-class of NSPPs—through the method of minimum contrast (Waagepetersen, 2007). Herein an approximate-likelihood approach is proposed, no claim is being made that this substantially improves parameter estimation. The method of minimum contrast is based on choosing a summary statistic which best reflects the data's structure. This, in order that the distance between the estimated and theoretical statistics should be minimised to estimate the parameters of the process. The Palm i.f returns the expected intensity of points at a distance r from a randomly chosen point, and is assumed to reflect the difference process (Tanaka et al., 2008). Although not available in a closed form expression for the Matérn process—as for the Thomas process—the log of the Palm likelihood is a simple affair to maximise using inbuilt `R` functions. The approach taken here only involves a one dimensional integral, drastically reducing the computational burden which would be associated with the necessity to integrate over all possible pairs of points.

2.6.2 Histopathology data

Beyond the estimation approach detailed in this chapter, there is of course scope to extend (and perhaps improve) the quantification of the spatial structure of the histopathology data discussed above. One might wish to consider other structures within the tissue, or extend the

analysis to 3D, data for which is becoming more available. The concept of the Palm likelihood approach—through the derivation of the Palm i.f—is hugely flexible, due the relative simple requirement of the derivation of the PDF, and hence CDF, of inter-sibling distances. This, of course is solely dependent on the complexity of the PDF. However, as described above this has the potential to describe complex spatial structures, hence, such methodology has the capacity for use in quantifying diverse tissue structures. It should also be noted that the flexibility of this approach—under the guise of Trace Contrast (TC) models (Stevenson et al., sion; Fewster et al., 2016)—has already seen use in an ecological setting.

2.6.3 Possible extensions and summary

In this Chapter Uniform dispersion of daughters around their parents are considered, an extension to the Gaussian dispersion of daughters considered by Stevenson et al. (*in submission*), who term this a Trace Contrast (TC) model. However, in order to consider other spatial dispersion of daughters around their parents, one simply requires the PDF—$f_y(r, \gamma)$—and hence the CDF—$F_y(r, \gamma)$—of the distance between two randomly selected siblings. Of course this—as in the Mateérn cluster process—may not lead to a tractable likelihood, yet may offer improvement over existing methods negating multiple integrations.

As far as the author is aware at the time of writing, applications of such models—fitted using the methodology discussed herein—have only been applied in the context of Trace Contrast models (which are an extension of the modified Thomas process). This by (Stevenson et al., sion) in the context of a simulated two-plane areal survey in the detection of cetaceans, and by Fewster et al. (2016) in the context of camera-trap detections of *Rattus rattus* (black rat) (here the detections were temporal in nature as opposed to spatial). In each case, the detections (traces) were taken as the daughters of a 2-dimensional or 1-dimensional NSPP respectively, where the parent density was assumed to reflect the true animal density. Yet, clearly (consider the application presented in this chapter) such models are not restricted to use with ecological data. Moreover, as demonstrated in this chapter, this methodology can be (relatively) easily extended to consider other distributions of daughters, which may better reflect the structure of the data.

The derivation of $f_y(r, \gamma)$, and hence $F_y(r, \gamma)$ (detailed above), one could envisage being extended to the multivariate case, enabling marked NSPPs to be considered. This conceptually would enable as one example multi-species interaction to be considered, which is of great interest in the study of population dynamics. Here, the points may represent animal location, and the marks their species, therefore the multivariate Palm i.f would return the expected intensity (conditional on the mark) of a marked point process at a discrete r from an arbitrarily chosen point.

Chapter 3, takes a similar approach to the one detailed here, where the Palm i.f intensity returns the expected intensity of a (newly derived) void process, a pattern with unusual gaps

(voids) given the general structure of the pattern. This model is yet again applied to the histopathology data yet only considers a single type of nuclei. Chapters following Chapter 3 take a slightly different approach in modelling point pattern data. In each case a LGCP is assumed, and fitted using the INLA - SPDE approach.

Chapter 3

An approximate-likelihood approach to estimate parameters of a void point process and its application in predicting cancer patient survival

3.1 Introduction

The focus of Chapter 2 was predominantly parameter estimation for NSPPs, a class of hierarchical cluster point process. Point processes are not restrained to only describe clustered point pattern data. Such data may exhibit other unusual spatial configurations. In particular this chapter proposes a new type of point process, herein called a void process. A void—in the context of point processes—may be thought of as a region which contains no points, where given the general structure of the data points may be expected to be. In some contexts a void may be thought of as a region where the characteristics of the points therein are atypical. For example, consider a cosmic void, which is a region devoid of bright galaxies (Kauffmann and Fairall, 1991). Parallels between the hierarchical structure of a NSPP (Chapter 2) and a void process can be drawn. Recall that clusters of a NSPP are assumed to be the result of unobserved parent points who each sire a number of observed daughters. The obstacle is in estimating the number of parents, having only observed the daughters. Herein, it is proposed that void processes are also hierarchical in nature. Similar to Chapter 2 this chapter names the observed points of the process daughters, and the unobserved points, parents. Again, the intensity of the unobserved parent points can be thought of in terms of a latent process driving the arrangement of daughters. In each case—that is both a NSPP and a void

process—the unobserved parent points, are themselves realisations of a homogeneous point process. These parent points induce structures in the point pattern, that is, either clusters of points or void spaces. Specifically, the unobserved parents dictate the structure of the observed pattern, resulting in this case in voids.

Cluster processes, have been widely considered in the point process literature. Perhaps the two most commonly considered are LGCPs (Møller et al., 1998) (considered in Parts III, IV, and V), and NSPPs (Neyman and Scott, 1958) (discussed in Chapter 2). Considered to a lesser extent, are point processes exhibiting void dynamics (i.e., areas devoid of points). The morphology of such processes are also inherent to the natural world. Such spatial structure is observed across many fields, (e.g., astronomy (Zeldovich et al., 1982), pedology (Brewer and Sleeman, 1960) and ecology (Stoyan and Penttinen, 2000)). For example, in the context of ecology, areas of the forest will naturally contain areas of fluctuating tree density; yet, there may also be areas, which, perhaps suffering from disease or storm damage do not contain any trees. Specifically the voids are created by unnatural means.

Characterising a point pattern which exhibits unusual empty areas (i.e., voids) must also consider the general structure of the pattern. Typically, algorithmic (Kauffmann and Fairall, 1991) or tessellation (Luchnikov et al., 1999) based approaches have been taken in the detection of voids. However, the aim of this chapter is not in identifying voids but in characterising the spatial structure of a void process. This through a vector of parameters, $\boldsymbol{\theta}$. However characterisation of a point process which adequately reflects the structure of interest is not always simple.

An approximate-likelihood approach is taken in this chapter to facilitate the estimation of $\boldsymbol{\theta}$; this, centred on the derivation of the Palm i.f (as in Chapter 2). No claim is made that the likelihood for a void process is intractable; yet, this chapter does promote using an approximate-likelihood approach. Section 3.1.1 presents methodology which is based on the inter-daughter distances of the pattern in the estimation of $\boldsymbol{\theta}$, through considering $L(\boldsymbol{\theta})$, termed the Palm likelihood (as in Chapter 2).

The ideas presented in this chapter are to be submitted to the *Annals of Applied Statistics* Jones-Todd et al. (*in submission*). The author of this thesis has in addition written an `R` package, `gapski`, not yet available on CRAN but available at `https://github.com/cmjt/gapski`. The functionality of this package is given in Section 3.2. An application to using the functionality of the package is given in Section 3.3 which considers histopathology data pertaining to colorectal cancer patient data, (as that considered in Chapter 2, and by Caie and Harrison (2016); Caie et al. (2014)). Here, it is assumed that the location of cell nuclei identified on histopathological slides are assumed to be realisations of a void process. This, in an attempt to ascertain if the morphology of cancerous tissue samples might aid in predicting patient survival.

3.1.1 Void process

Point patterns whose morphology exhibit unexpected areas devoid of points, are considered to be realisations of what are herein named void processes. Little headway has been made into such processes. Similarities may be made between void processes and regular processes. Processes giving rise to regular point patterns, such as Strauss processes[1], are generalisations of the homogeneous Poisson process. Here the points are no longer independently distributed in space; in each case points are repelled by one another, this repulsion is constant within a fixed interaction radius, r_0, around each point. The strength of this interaction ranges from no interaction—equivalent to a homogeneous Poisson process—to outright inhibition within radius, r_0, around each point—where no points are observed within distance r_0 from any observed point. Along a similar vein, another class of interaction process, hardcore processes, cannot contain points which are closer than a distance, say r_0, apart. One such class of processes are hardcore Matérn processes, which are the result of some dependent thinning operation (Matérn, 1986). Here points do not exist—due to some thinning operation—within a certain radius, r_0, of each point in the observed (thinned) point process. Inhibition and hardcore processes, by definition, describe regularity in point patterns due to some inhibitive interaction (Stoyan, 1988). Despite their similarity to these, void processes are in fact not regular; realisations of void processes exhibit some areas devoid of points but not others.

As in the case of NSPPs, discussed in Chapter 2 above, void processes are hierarchical in nature. That is, there are assumed to be unobserved points which somehow dictate the structure of the observed pattern. As opposed to NSPPs, where (unobserved) parent points sire (observed) daughter points which form clusters, here (unobserved) parent points result in voids amongst their (observed) daughters. Specifically, assuming a homogeneous pattern of points, parents expunge every daughter in a sphere of radius R centred at the parent's location. Hence, the remaining daughters are free from their parent's influence and are realisations of a void process. Two realisations of a void process[2] are shown in Figure 3.1, in both cases (as the data were simulated) the (otherwise unobserved) parent locations are indicated by grey crosses. Moreover, the dotted circles encircling each parent cross illustrate the voids. These patterns are specifically simulated[3] to illustrate how void identification is almost possible by eye when considering a dense pattern (left hand plot) and impossible

[1] a special case of a Gibbs process Illian et al. (2008)

[2] Functionality of the `R` package `gapski` detailed in Section 3.2.1 will enable the reader to recreate these patterns.

[3] running (functionality of the `R` package `gapski`)

```
set.seed(222)
sim<-sim.gap(pars=c(D=10,lambda=300,R=0.075),plot.points=TRUE)
set.seed(4321)
sim<-sim.gap(pars=c(D=5,lambda=30,R=0.1),plot.points=TRUE)
```

will result in the patterns shown in the left and right plots of Figure 3.1 respectively

when considering a much sparser pattern (right hand plot). Others have demonstrated that on some occasions it is possible to identify voids by eye. An example shown in Illian et al. (2008, p. 258) visually identifies storm damage in a plot of trees through the use of a relatively simple 4-neighbour graph technique. However, for sparser patterns, and patterns with nearby voids this becomes much more difficult. It should be noted, however, that the main focus of this chapter is not in the identification of voids, but in estimating their density. The concept of not identifying voids, is similar to the approach taken in Chapter 2 where no daughter was assigned cluster identity. In both instances it is not this classification (i.e., identifying voids or cluster assignation) which is of interest, but the density of these structures (i.e., parent density).

A void process is defined by both the distribution of the number of parents, and the distance between two independent daughters. The number of parents, is simply a random variable which follows a Poisson distribution. The distribution of the distance between two independent daughters depends on the expunging radius of the parents. A void process, can be thought of as an example of a dependently thinned point process (Illian et al., 2008, p. 365). Here D gives the intensity of the unobserved parents, and λ the intensity of observed daughters (i.e., outwith the voids, having escaped their parents). The parameters of a void process are thus, $\boldsymbol{\theta} = (D, \mathrm{R}, \lambda)$.

In order to estimate these parameters a similar approach as detailed in Chapter 2 is taken. Again, the approach taken is centred on deriving the Palm i.f for the process, after which the Palm likelihood, $L(\boldsymbol{\theta}; \mathbf{r})$ (discussed in Section 2.3.2), is derived. Whence following the methodology outlined in Chapter 2, the estimator for $\boldsymbol{\theta} = (D, \mathrm{R}, \lambda)$ is given by,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \mathbf{r}),$$

which is evaluated through the numerical maximisation of $\log(L(\boldsymbol{\theta}; \mathbf{r}))$ with respect to $\boldsymbol{\theta}$.

The remainder of this section derives the Palm i.f for a void process, and therefore the Palm likelihood. Thereafter, functionality of the `R` package `gapski`[4] in simulating and fitting void processes is discussed. Moreover, Section 3.3.1 considers an application pertaining to the colorectal cancer patient dataset, discussed in Chapter 2 above and detailed in Caie and Harrison (2016), and Caie et al. (2014).

Palm intensity of a void process

The Palm i.f of a void process—as in the case of any other point process (see Section 2.3)—gives the expected intensity of observed points at a distance r from another arbitrary chosen point. The Palm i.f, $\lambda_0(r)$, for a void process, following a similar formulation of Illian et al. (2008, p. 220) is given by,

$$\lambda_0(r) = \lambda\, p_s(r), \tag{3.1}$$

[4]available from `https://github.com/cmjt/gapski`

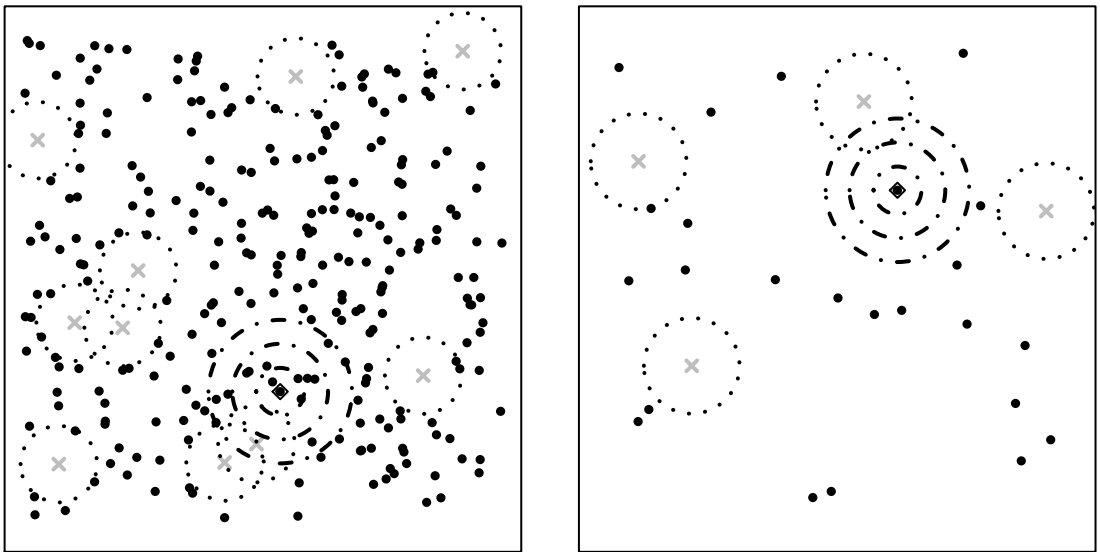


Figure 3.1: Two simulated void processes in the unit square. With (unobserved) parent points (crosses) having expunged daughters within a distance R so that only the observed daughters remain (dots). The daughters are generated by a homogeneous Poisson process with densities $\lambda = 300$ and $\lambda = 30$ (left and right) respectively. The number of parents simulated are IID following a Poisson distribution with expectation $D = 10$ and $D = 5$ respectively. The distance beyond which daughters are safe from their parents are R = 0.075 and R = 0.1. The dotted circles indicate the simulated voids. Noting an arbitrarily chosen daughter (encircled triangle) the distance between her and her sister is clearly related to the proximity of a parent. The dashed circles show how the intensity of observed daughters changes for different distances r away from an observed daughter. Clearly for higher λ (left hand plot) it is easier to identify voids by eye, however for sparser patterns it becomes much harder to identify such areas.

where $p_s(r)$ is the probability that an arbitrary point of the process has no parent within some distance R. Thence, $p_d(r) = 1 - p_s(r)$, is the probability that a possible point is within expunging distance of a parent. Thus, given an observed daughter a nearby point is also likely to be outwith expunging distance. Hence, the Palm i.f is large for small r decaying to some asymptote for large r.

The dash-dot spheres encircling the arbitrarily chosen observed point in Figure 3.1 will not contain an (unobserved) parent at a distance less than R away. Thence, the probability that potential point at distance r from the observed point is also observed is related to the volume of intersection of the hyperspheres of radius R encircling the observed point and the potential point. The geometry of this concept is illustrated by Figure 3.2. Here, the closed plotting character represents an observed point, the open plotting character represents a possible point. As there cannot exist a parent within distance R of the observed point, the intersection, I(r), is the area within which any possible point is safe. This, is therefore the volume within a distance R of the observed point, where no parent can be observed. The remaining volume, $A(r)$, of the hypersphere encircling the potential point is the only region that may contain a parent. As the distribution of parents is assumed to be a Poisson point process, their number follows a Poisson distribution with expectation $D\,A(r)$. Hence, $p_s(r)$

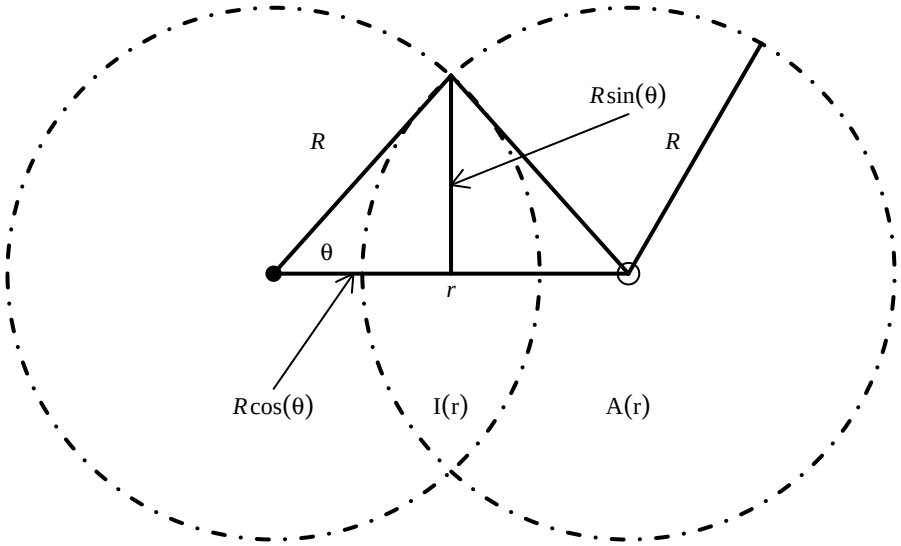


Figure 3.2: Figure showing the intersection of two circles, of radius R. The closed plotting character represents an observed point, the open plotting character represents a possible point. There cannot exist a parent within distance R of the observed point, therefore the intersection $I(r)$ is the area within which any possible point is "safe". In general, although only illustrated in 2D here, to ascertain the intersection between two hyperspheres of common radius R the radius of the hyper-spherical caps, $\mathrm{R}\sin(\theta)$, and the height of the hyperspherical caps $\mathrm{R}\cos(\theta)$ is required, where θ is the colatitude angle. Thus, the volume of intersection only depends on the radii of the hyperspheres as well as the distance between their centres r, see Equation (3.2).

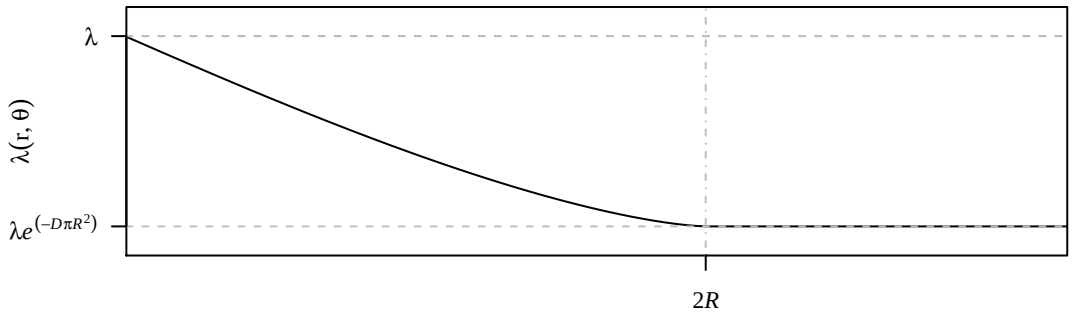


Figure 3.3: The functional form of the Palm i.f for the void process in 2 dimensions. The horizontal asymptote is $\lambda\exp(-D\pi\mathrm{R}^2)$. This is the value that $\lambda_0(r)$ decays to for values of $r \geq 2\,\mathrm{R}$. The Palm i.f for $r = 0$ is simply λ, as for $r = 0$, $g(r) = 1$ (Equation (3.3)), thus $\exp(\cdot) = 1$.

is given by,

$$p_s(r) = \exp(-D\,A(r)),$$

where $A(r)$ can be determined through the use of elementary geometry (see Figure 3.2).

The volume of intersection between two d-dimensional spheres with a common radius R is derived as in Li (2011). To calculate $I(r)$ (shown in Figure 3.2) the integrand of the volume of a $\mathrm{d}-1$ sphere of radius $\mathrm{R}\sin(\theta)$ with height $\mathrm{R}\cos(\theta)$ must be derived. As the hyperspheres are of common radius, R, $I(r)$ is simply just twice this volume. So, letting $I^{\mathrm{d}}(r)$ be this

volume of intersection in d-dimensions, then,

$$\begin{aligned}
I^{\mathrm{d}}(r) &= 2 \times \int_0^{\phi} v^{\mathrm{d}-1}(\mathrm{R}\sin\theta)\,\mathrm{dR}\cos\theta, \\
&= 2 \times \int_0^{\phi} v^{\mathrm{d}-1}(\mathrm{R}\sin\theta)\mathrm{R}\sin\theta\,\mathrm{d}\theta, \\
&= 2 \times \frac{\pi^{\frac{\mathrm{d}-1}{2}}}{\Gamma(\frac{\mathrm{d}-1}{2}+1)}\,\mathrm{R}^{\mathrm{d}} \int_o^{\phi} \sin^{\mathrm{d}}\theta\,\mathrm{d}\theta, \\
&= 2 \times \frac{\pi^{\frac{\mathrm{d}-1}{2}}}{\Gamma(\frac{\mathrm{d}-1}{2}+1)}\,\mathrm{R}^{\mathrm{d}}\,\mathrm{J}_d(\phi), \\
&= \frac{\pi^{\frac{\mathrm{d}-1}{2}}}{\Gamma(\frac{\mathrm{d}-1}{2}+1)}\,\mathrm{R}^{\mathrm{d}}\,B\left(\frac{\mathrm{d}+1}{2},\frac{1}{2}\right)\,\mathrm{I}\left(\sin^2\phi;\frac{\mathrm{d}+1}{2},\frac{1}{2}\right), \\
&= v^{\mathrm{d}}(\mathrm{R})\,\mathrm{I}\left(\sin^2\phi;\frac{\mathrm{d}+1}{2},\frac{1}{2}\right), \\
&= v^{\mathrm{d}}(\mathrm{R})\,\mathrm{I}\left(1-\left(\frac{r}{2\mathrm{R}}\right)^2;\frac{\mathrm{d}+1}{2},\frac{1}{2}\right),
\end{aligned} \tag{3.2}$$

noting that $B(a,b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}$ and $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. Here, the integral $\mathrm{J}_d(\phi)$ is given by,

$$\begin{aligned}
\mathrm{J}_d(\phi) &= \int_0^{\phi} \sin^d\theta\,\mathrm{d}\theta \\
&\quad \frac{1}{2}B\left(\sin^2\phi,\frac{\mathrm{d}+1}{2},\frac{1}{2}\right),
\end{aligned}$$

and $\mathrm{I}(z;a,b) = \frac{B(z;a,b)}{B(a,b)}$ is the regularised Beta function. Here , $v_{\mathrm{d}}(\mathrm{R})$ is the d-dimensional volume of a hypersphere of radius R. The above derivation is achieved through noting that,

$$(\mathrm{R}\cos\theta)^2 + (\mathrm{R}\sin\theta)^2 = \mathrm{R}^2 \quad \rightarrow \quad (\sin\theta)^2 = 1 - (\cos\theta)^2,$$

and that through using the cosine rule,

$$\cos\theta = \frac{r^2 + \mathrm{R}^2 - \mathrm{R}^2}{2\,r\,\mathrm{R}} = \frac{r}{2\,\mathrm{R}}.$$

The Palm i.f given by Equation (3.1) is thus,

$$\begin{aligned}
\lambda_0(r) &= \lambda \exp\left(-D\,v^{\mathrm{d}}(\mathrm{R})\left[1-\mathrm{I}\left(1-\left(\frac{r}{2\mathrm{R}}\right)^2;\frac{\mathrm{d}+1}{2},\frac{1}{2}\right)\right]\right), \\
&= \lambda \exp\left(-D\,v^{\mathrm{d}}(\mathrm{R})\left[1-\mathrm{F}_{g(r)}\left(\frac{\mathrm{d}+1}{2},\frac{1}{2}\right)\right]\right),
\end{aligned} \tag{3.3}$$

where $g(r) = 1 - \left(\frac{r}{2\mathrm{R}}\right)^2$, $\Gamma(\cdot)$ represents the Gamma function, and $\mathrm{F}_{g(r)}(\cdot,\cdot)$ is the CDF of the Beta distribution. This Palm i.f is shown in Figure 3.3. This function is piece-wise continuous with two sub-domains, $(0, 2\mathrm{R}]$, and $[2\mathrm{R}, \infty)$. This as, at the value $r = 2\mathrm{R}$ the volume of intersection between the spheres encircling an observed daughter and a potential point of radius R is zero. Thus, the contribution from the CDF of the Beta distribution to

$\lambda_o(r)$ is zero. Further properties of this Palm i.f are illustrated in Figure 3.3. In particular, when $r = 0$ the pair c.f, $g(r) = 1$ which leads to $\mathrm{F}_1(\cdot,\cdot) = 1$ and thus $\lambda_0(0) = \lambda$, the initial spike of the Palm i.f. In addition when $r = 2\mathrm{R}$ the pair c.f, $g(r) = 0$ leading to $\mathrm{F}_0(\cdot,\cdot) = 0$ and then $\lambda_0(0) = \lambda \exp(-D\, v_{\mathrm{d}}(\mathrm{R}))$ (due to the properties of the CDF). This is the horizontal asymptote of the Palm i.f.

Estimating parameters of the void process

Following the methodology described in Section 2.2.1 for the NSPP, estimation of $\boldsymbol{\theta} = (D, \mathrm{R}, \lambda)$ for the void process is again achieved through maximising the log of the Palm likelihood. That is, the d-dimensional differences of the void process are assumed to be realisations of an inhomogeneous Poisson process with intensity given by Equation (3.3). Hence the Palm likelihood is given by,

$$L(\boldsymbol{\theta}; (r)) = \left(\prod n\, \lambda_o(r; \boldsymbol{\theta})\right) \exp\left(-n \int_0^t \lambda_o(r; \boldsymbol{\theta})\, s^{\mathrm{d}}(r) \mathrm{d}r\right), \tag{3.4}$$

where, $s^{\mathrm{d}}(r)$ is the volume of revolution. The integral is a volume of d-dimensions around the intensity axis, it is intractable as it contains the CDF of a Beta distribution. However it can be simplified, the integrand in Equation (3.4), is simplified as follows,

$$\begin{aligned}
\int_0^t \lambda_o(r; \boldsymbol{\theta})\, s^{\mathrm{d}}(r) \mathrm{d}r &= \int_0^t \lambda \exp\left(-D\, v^{\mathrm{d}}(\mathrm{R}) \left[1 - \mathrm{F}_{g(r)}\left(\frac{\mathrm{d}+1}{2}, \frac{1}{2}\right)\right]\right) s^{\mathrm{d}}(r)\, \mathrm{d}r, \\
&= \lambda \int_0^t \exp\left(-D\, v^{\mathrm{d}}\right) \exp\left(D\, v^{\mathrm{d}} \mathrm{F}_{g(r)}\left(\frac{\mathrm{d}+1}{2}, \frac{1}{2}\right)\right) s^{\mathrm{d}}(r)\, \mathrm{d}r, \\
&= \frac{\lambda\, \mathrm{d}\, \pi^{\frac{\mathrm{d}}{2}} \exp\left(-D\, v^{\mathrm{d}}\right)}{\Gamma\left(\frac{\mathrm{d}}{2}+1\right)} \int_0^t \exp\left(D\, v^{\mathrm{d}} \mathrm{F}_{g(r)}\left(\frac{\mathrm{d}+1}{2}, \frac{1}{2}\right)\right) r^{\mathrm{d}-1}\, \mathrm{d}r,
\end{aligned}$$

where $g(r) = 1 - \left(\frac{r}{2\mathrm{R}}\right)^2$, $\Gamma(\cdot)$ represents the Gamma function, and $\mathrm{F}_{g(r)}(\cdot,\cdot)$ is the CDF of the Beta distribution.

3.2 The `gapski` package

This section focuses on the implementation of an approximate-likelihood method in order to estimate parameters of a void process. Specifically through the maximisation of the (approximate) likelihood given by Equation (3.4). This is achieved through utilising functionality of the `R` package `gapski`, developed by the author of this thesis. It should be noted that most of the functionality given below is based on an early version of that package. The reminder of this section introduces, and describes functionality of the `gapski` package, along with providing examples for various functions found within the package[5].

[5]This `R` package is available at `https://github.com/cmjt/gapski`

3.2.1 Simulating a void process–the `sim.gap()` function

The `gapski` package contains two sets of simulated data, which can be found in the exported objects `example.void.1D` and `example.void.2D`[6]. Each provide realisations of a void process in one- and two- dimensions, respectively. These realisations were generated by the function `sim.gap()`, which simulates realisations of a void process. The main arguments of the function are:

- `pars`: A named vector of model parameter values, $\boldsymbol{\theta}$. Required parameters are D, the density of gaps in the pattern; λ the density of all points in the domain, and R the radius of the spherical gaps in the pattern.
- `d`: A numeric value specifying the number of dimensions the point process is to be simulated in, by default this is 2.
- `lims`: A matrix with d rows giving the limits of the hyper-rectangular observation window.

Additional arguments allow for either the simulated point pattern or the empirical Palm i.f to be plotted, as well as the simulated parent locations to be returned (if this argument is chosen a list is returned). As in Section 2.4 above the simulated locations are subject to periodic boundary conditions. To simulate a realisation on the void process with $D = 10$, $\lambda = 300$, and $\mathrm{R} = 0.075$ in the unit square, the following is used[7]

```
points<-sim.gap(pars=c(D=10,lambda=300,R=0.075))
```

The object returned is a matrix[8] with the number of rows pertaining to the number of simulated daughters and 2 columns, providing the locations of the simulated daughters.

3.2.2 Fitting a void process–the `fit.gap()` function

Estimation of $\boldsymbol{\theta}$ for a void process is achieved via the `fit.gap()` function. Maximisation of the log of the Palm likelihood, (Equation (3.4)) is done numerically through use of the `optim()` function using the `'L-BFGS-B'` method. The function has one equivalent argument to the `sim.gap()` function, that is, `lims`. The other main arguments are:

- `points`: a matrix containing the observed points–the object returned by `sim.gap()` is appropriate[9].

[6] This realisation is the same as the object `points`, discussed in this section, that is a void process with parameters $\boldsymbol{\theta} = (10, 0.075, 300)$

[7] running `set.seed(222)` prior to executing the code will result in the point pattern plotted in the left hand plot of Figure 3.1.

[8] unless `return.aunts=TRUE` is chosen–in which case a named list of matrices is returned.

[9] unless the aunt locations are returned

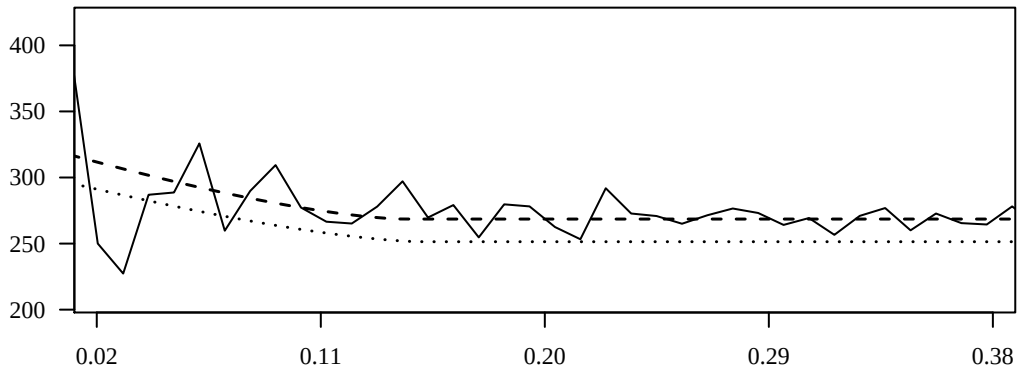


Figure 3.4: Plot showing the empirical palm intensity (solid line) of the simulated object `points` above, along with the fitted palm intensity (dashed line), and the actual palm intensity (dotted line).

- `trunc`: the truncation distance, that is, a distance beyond which the intersection between the intersection of hyperspheres of common radius R centred at an observed niece and a possible point is thought to be zero.
- `D.sv`: the start value for D is required for the optimisation algorithm.
- `D.bounds`: a numeric vector of length two containing lower and upper bounds for the parameter D.

A void process is thus fitted to the realisation above through running the following code.

```
fit.gap<-fit.gap(points = points,lims = rbind(c(0,1),c(0,1)),trunc = 0.5,
                 D.sv = 10,D.bounds=c(7,13))
```

This returns an object providing information about the fitted model which can be extracted through utility functions. In particular, `summary(fit.gap)` will provide a model summary, `coef(fit.gap)` will extract the estimated model parameters, and `plot(fit.gap)` will plot the fitted palm intensity function.

In order to illustrate estimator performance 1000 points patterns were simulated with parameters $D = 10$, R $= 0.075$, and $\lambda = 300$ (the same as the object `points` above). Figure 3.4 shows the empirical Palm i.f (solid line), the fitted Palm i.f (dashed line), and the actual Palm i.f (dotted line) of one such point pattern (in fact the one shown in the left hand plot of Figure 3.1). Figure 3.5 and Table 3.1 show the estimator performance of the approximate-likelihood approach discussed herein. And as can be seen there is a slight positive bias for the void density parameter.

3.2.3 Variance estimation–the `boot.gap()` function

Following the methodology discussed in Chapter 2 above, variance estimation is carried out using a parametric bootstrap. This is implemented by the function `boot.gap()`. The main

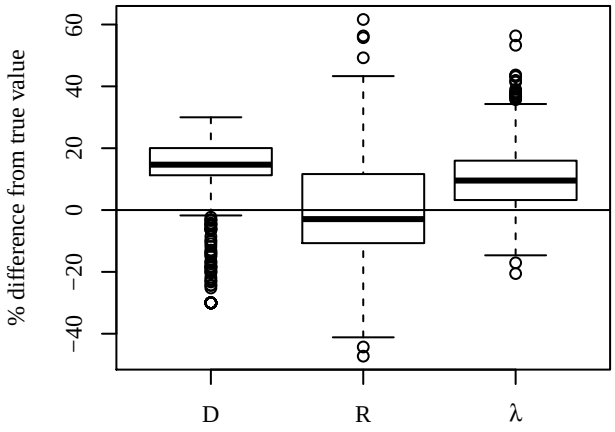


Figure 3.5: Boxplots of estimates showing the percentage difference from the true parameter values used to simulate the void point pattern data with parameters $D = 10$, R = 0.075, and $\lambda = 300$.

	$Bias(\%)$	$RMSE(\%)$
D	13.68	3.31
R	1.48	2.91
λ	9.99	1.97

Table 3.1: Estimated biases, and root mean square errors from the simulation studies given as percentages of the true parameter values used to simulate the data.

arguments of `boot.gap()` are simply the fitted model, supplied as the first argument, and N, the number of bootstrap resamples. Thus, a 1000 resample bootstrap can be carried out through running the following `R` code,

```
boot.gap.fit<-boot.gap(fit.gap,N=1000)
```

Through the use of utility functions, in particular `summary()` and `confint()`, the parameters, their standard errors, and confidence intervals can be returned by using utility functions in R such as the following.

```
summary(boot.gap.fit)
##                coefs          se
## R         0.07146377  0.01273723
## D        11.29713116  1.20050542
## lambda 321.92438383 29.59613256
## attr(,"class")
## [1] "summary.gapski" "matrix"
confint(boot.gap.fit)
##                2.5 %     97.5 %
## R         0.05347096   0.101791
## D         7.00000000  13.000000
## lambda 277.10224198 395.677596
```

3.2.4 Void detection–the `voids.gap()` function

Once estimation of $\boldsymbol{\theta}$ has been carried out using the functions described above, use of the `voids.gap()` function will aid in a somewhat rudimentary detection of voids in the point pattern. This (rather heuristic) procedure is purely algorithmic and depends on a Voronoi-Delauney tessellation approach. This, coupled with the estimated parent density and void radius enables the most likely void locations to be identified. The only argument to `voids.gap()` is the model object returned from running `fit.gap()`. Calling the function on the model object `fit.gap` above results in the left hand plot of Figure 3.6. Here, the asterisk show the true location of the simulated parents. The proposed void finding procedure does miss some parents. This probably due to a reduced amount of information resulting from parents near the edge of the pattern. However, as can be seen the procedure does manage to identify seven of the parent locations correctly.

Recall the previously mentioned data discussed by Illian et al. (2008, p. 257) referring to locations of trees in a Sitka spruce forest. They suggest an approach based on a 4-neighbour to detect areas of storm damage. The approach proposed herein uses the parameter estimates of the void process fit in conjunction with a Voronoi-Delauney tessellation approach to identify voids. Upon fitting a void process model to the Sitka spruce data mentioned above, the right hand plot of Figure 3.6 shows the results of this approach in identifying voids in the pattern (illustrated by the grey circles). It should be noted that these indicated areas were identified in Illian et al. (2008) (upon field investigation) as being regions of the storm damage.

3.3 Application to histopathology data

Recall that a void, in the context of spatial point patterns, is a region that contains no point where, given the general structure of the data, points may be expected. This section assumes that each of the separate point patterns formed by tumour and stroma nuclei (the data discussed in Chapter 2) are realisations of void processes. An illustration of this data is given in Figure 2.4. The void process in this setting is assumed to characterise the gaps in the distribution of nuclei. In this way, the void process describing the tumour cell point pattern reflects the patterning of stroma cells and vice versa; note the void processes also reflect the less frequently occurring regions of necrosis/lumen and areas of no tissue.

The void (parent) density D, does not necessarily refer to a physical construct, rather refers to the morphology of the tissue. Herein, it is assumed that(i) parent points are realisations of a homogeneous Poisson point process, with expected intensity, D; (ii) no point is observed within a distance R of an unobserved parent; and (iii) the intensity of daughters is given by λ. The following section investigates whether the spatial distribution of tumour cells alone can aid in discriminating between patient survival outcomes.

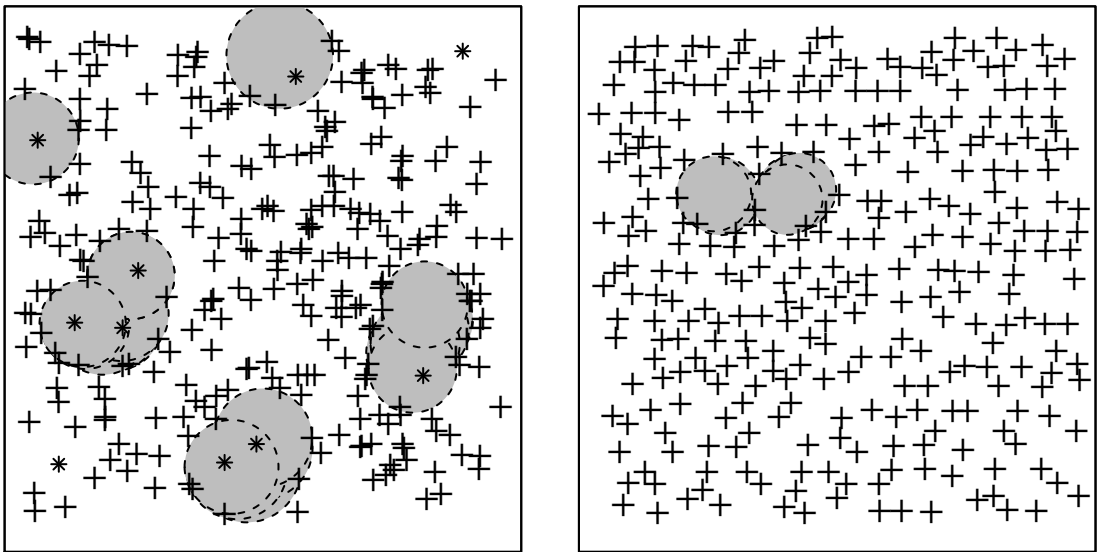


Figure 3.6: Resulting plot from calling the `voids.gap()` function. Left: indicating locations of potential voids in the simulated void process shown in the left hand plot of Figure 3.1, asterisks show the true locations of the simulated parents. Due to a reduced amount of information resulting from parents near the edge of the data, the rather heuristic void finding procedure does miss these parents. However, as can be seen the procedure does manage to identify 7 of the parent locations correctly. Right: indicating areas of potential storm damage in a Sitka spruce forest, the indicated areas correspond to those identified by Illian et al. (2008, p. 258) who claim that the identified regions pertain to actual areas of storm damage upon field investigation.

3.3.1 Results

Following estimation of $\boldsymbol{\theta} = (\mathrm{R}, D, \gamma)$ (for each of the stroma and tumour patterns) a hierarchical bootstrap procedure using 1000 simulations was carried out. This technique was also employed in Chapter 2; this as, individual patterns are not assumed to be independent, and thus the bootstrapped samples are devised to have the same dependence structure as the original sample. Table 3.2 contains the medians and bootstrapped quantiles of the distributions of the parent densities, $\hat{D}_{\cdot}^{\cdot}$, and void radii, $\hat{\mathrm{R}}_{\cdot}^{\cdot}$, pertaining to each point pattern (subscripts t and s refer to tumour and stroma patterns respectively). Here the superscripts d and a refer to whether the patients were dead or alive at follow-up respectively. The results in Table 3.2 reveal three findings in reference to the tumour patterns: (i) in the distribution of void densities there is no overlap between the patients who lived and the patients who died, (ii) the values of the parent densities are lower for patients who lived than for patients who died, (iii) the estimated void radii pertaining to patients who lived and died are similar. When considering the stroma patterns, there are no clear differences between patient groups for density or radii.

In order to assess the predictive power of each parameter (classifier) of each void process discussed above (i.e., $D_{\cdot}^{\cdot}$, void density of a void process respectively for tumour and stroma patterns (subscripts t and s respectively)) in predicting patient outcome ROC curves were used. Specifically, the means of each classifier were calculated per patient for use in a logistic

		Quantile of Order	
	Median	2.5%	97.5%
$\hat{D}_t^a$	8.6337	7.7222	9.3917
$\hat{D}_t^d$	15.4191	12.7417	18.1622
$\hat{R}_t^a$	0.2142	0.1977	0.2288
$\hat{R}_t^d$	0.2122	0.1858	0.2417
$\hat{D}_s^a$	6.0304	5.3665	7.9776
$\hat{D}_s^d$	5.4828	5.0402	6.5505
$\hat{R}_s^a$	0.2097	0.1929	0.2229
$\hat{R}_s^d$	0.2149	0.1976	0.2296

Table 3.2: The medians and bootstrapped quantiles (based on 1000 simulations) of the distributions of the parent densities and void radii for each of the point patterns formed by both tumour and stroma cells (subscript t and s respectively) of either patients who lived or died (superscripts a and d respectively) assuming a void process.

regression (in predicting patient outcome). Figure 3.7 shows the ROC curves for each classifier along with the associated area under the curve (AUC) and leave one out cross validation (CV) scores. Notably, void density of the tumour pattern is the best predictor for patient outcome (AUC of 0.84 and CV of 0.19). Thus, when considering the AUC scores, the daughter density of the assumed Thomas process (see Chapter 2) for the stroma patterns is considered the second best predictor.

The model assumes that a void process adequately describes the spatial structure of the cells. To informally asses the validity of this assumption the spatial patterning of the observed data and data simulated with the same estimated parameter values from the model are compared, in a similar way to the NSPP processes in Section 2.5.1. This comparison is again accomplished using the empty space function. Figure 3.8 shows the empty space functions for the void process for the slide shown in Figure 2.4. Each solid line represents the empty space function for the fitted model; the dotted lines are the empty space functions estimated for patterns simulated with the respective estimated parameters. Based on Figure 3.8 we can see that a void process seems an adequate assumption to make for each cell type if measured by the empty space function. As noted in Section 2.5.1 comparing the empty space function of both the tumour and stroma cell patterns to that of the homogeneous Poisson process (thick dashed lines) it seems that at short distances the cells exhibit some regularity, whereas only at longer distances does the pattern exhibit clustering. This is expected, as in both cases the points represent cell nuclei, and due to the size of the cells at short distances no other cells (points) can exist. Again it should also be noted that the stroma cell pattern more quickly approaches that of a homogeneous Poisson process. From Figure 3.8 this is the case at a distance of 3% of the width and height of the slide ($r = 0.03$); this seems reasonable from the pattern of stroma cells shown in Figure 2.4. This also is reflected by the estimated

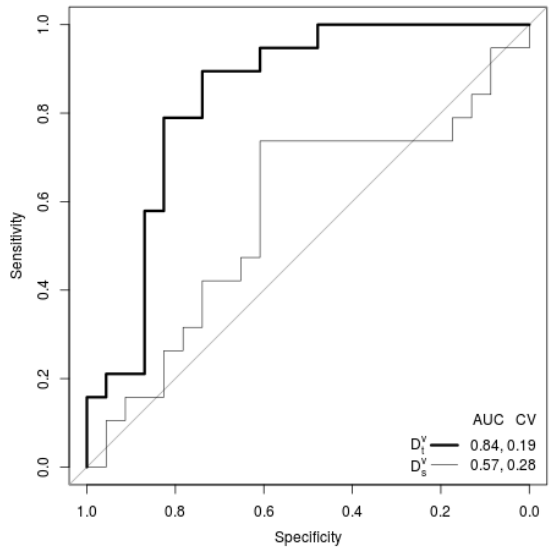


Figure 3.7: Plot of the ROC curves in predicting patient outcome. Legend includes the AUC (area under the curve) score for each curve and the CV (leave one out cross validation) score for each classifier (i.e., $D_{\cdot}^{v}$, void densities for a void process))

lower void density for stroma cells (Table 3.2), which would lead to a pattern with fewer gaps.

This section's main focus was on void process analysis of gaps in either tumour or stroma cells, which reflects the stroma or tumour cell patterning respectively. The tumour cell void patterning is found to be the most discriminatory, and has the best predictive power. This implies that stroma patterning is a better indicator of patient outcome than tumour patterning. This section chiefly presented a novel application of spatial statistics to capture the spatial arrangement of cells using a void process, thus illustrating the value of analytical methods to understand the spatial structure of cancer. Most notably, it is the predictive ability of the classifiers mentioned above which is of interest. It was demonstrated that from the spatial structure of cancerous tissue one can quantitatively infer a patient's outcome.

3.4 Discussion

In the above chapter methodology similar to that discussed in Chapter 2 is proposed. This in order to estimate parameters of the void process derived herein. The estimation procedure—as in Chapter 2—is centred around the derivation of the Palm i.f, and thus the maximisation of the Palm likelihood. The proposed void process is assumed to be a hierarchical process where Poisson distributed unobserved parent points—with density D—represent hyper-spherical voids of radius R, within which no daughter point can be observed. The parameter vector of the derived void process is thus given by $\boldsymbol{\theta} = (D, \mathrm{R}, \lambda)$, where λ gives the density of daughters (D and R are as described above).

The shape of the Palm i.f for the void process is similar to that derived for the Matérn

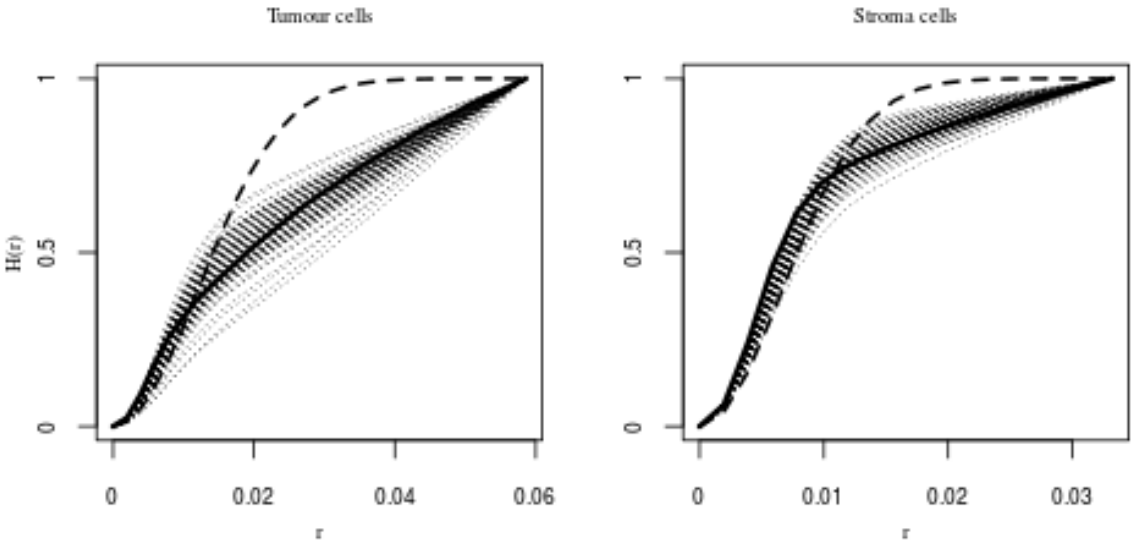


Figure 3.8: Figure showing for one patient's slide (Figure 2.4) the fitted empty space function, $H(r)$, by the solid line, and those for 100 simulated patterns (simulated using the estimated parameter values for each process), dotted lines. The thick dashed line is the theoretical empty space function for a homogeneous Poisson process. The distance r represent the percentage of the slide's width and height scaled to be $\in [0, 1]$.

cluster process in the previous chapter, that is both are piece-wise continuous functions with two sub-domains, $(0, 2\mathrm{R}]$, and $[2\mathrm{R}, \infty)$. The similarity in shape is due to the geometry of intersecting hyper-spheres, the two-dimensional case is illustrated in Figure 3.2. Specifically, at a distance 2R, for both the Matérn cluster process, and the void process this is the distance beyond which circles of radius R centred at two arbitrary points cannot intersect.

3.4.1 Parameter estimation for a void process

Apart from algorithmic void detection methodology, little thought has been given to what are herein termed void processes. Literature that does discuss voids is typically restrained to astrology (Zeldovich et al., 1982). However, some mention can be found in the fields of pedology (Brewer and Sleeman, 1960) and ecology (Stoyan and Penttinen, 2000). In these cases generally the interest is in the detection of voids; typically, through utilising an algorithmic procedure (Kauffmann and Fairall, 1991) or through the use of a tessellation approach (Luchnikov et al., 1999). In this chapter an approximate-likelihood approach is taken involving the maximisation of the Palm likelihood—a generalisation of the Poisson likelihood—in order that the density of the voids (and other associated parameters) may be estimated. This methodology, is applied in the consideration of the spatial structure of cancerous tissue samples at a cellular level. Using this technique facilitates (to some extent) the quantification of the cellular morphology. The information contained in spatial structure of the tissue aids in inferring the severity of tumour. Being able to quantify this spatial structure is fundamental to predict patient survival. It is generally the pathologists job to visually assess the mor-

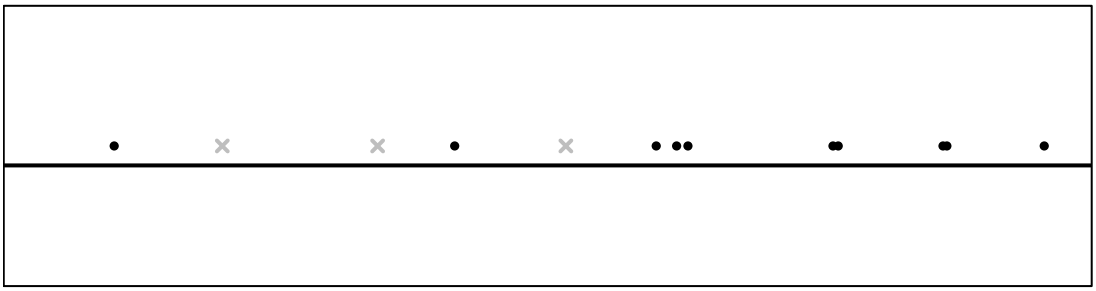


Figure 3.9: Simulated 1D conceptualised observation process, using functionality of the `R` package `gapski`, with parameter values $\boldsymbol{\theta} = (5, 0.075, 20)$, purely for illustration purposes. Here the grey crosses (parents) illustrate locations where there existed an animal, yet due to the failure of the observer are not recorded.

phology of the tissue samples and gauge the tumour's severity. Despite their expertise this assessment of the tissue is still sensitive to subjectivity. No claim is being made that the point process methodology—the fitting of void process models—aids in understanding the biological processes generating the cellular morphology. Yet, using the methodology presented in both this and the previous chapter demonstrated that encapsulating the spatial complexity inherent in histopathology data is possible from a point process point of view. Moreover, this spatial structure contains information, which may—as it has in this application—act as an indicator of patient survival. Notably, void density of the tumour pattern is found to be the best predictor for patient outcome.

No extension beyond the void process discussed is proposed in this thesis. However, one might imagine many; for example, the process could be extended to describe the non-homogeneous distribution of parents, or indeed daughters. This would in all probability lead to a rather involved Palm i.f, and hence the maximisation of the Palm likelihood would become very computationally expensive. The usefulness of relaxing the homogeneity assumption, for either the (unobserved) parent or (observed) daughter points, is not clear in practice. One conception relates to an observation process where the aim would be to distinguish between true and false zeros. Figure 3.9 illustrates a single line transect; here, each plotting character represents the locations of the species of interest (black filled circles—those observed, grey crosses—those missed by the observer). This figure serves as an example of the concept given above. That is, how to distinguish between the gaps induced by there being no animal present, and those induced by an animal being missed (grey crosses). However, in this situation it may be inappropriate to assume that animal distribution along the transect is homogeneous; hence, deriving the Palm i.f would be far more involved than that considered above. No further thought has been given to this concept beyond the brief remit discussed here and illustrated by Figure 3.9. Thus, it is left to the reader, and further consideration of the author to contemplate, and apply such methodology.

The use of a Palm i.f facilitates applying an approximate-likelihood technique in the fitting

of point process models as has been illustrated by the preceding chapters. However, other techniques may be employed to fit such models; the following chapters focus on exploring some such ways of fitting point process models. In particular, the focus is fitting LGCPs models which assume some nonstandard latent structure in a computationally efficient way (full details of the author's meaning of nonstandard are given in the following chapters).

Part III

Accounting for spatial dependency in real world data

Chapter 4

Modelling point pattern data using a log-Gaussian Cox process

4.1 Introduction

In the context of point pattern data, it is the spatial structure that is of interest. This spatial structure is epitomised by the intensity, points per unit volume. Allowing this intensity to vary spatially leads to an inhomogeneous point process. In the context of an inhomogeneous Poisson process the intensity is given by some location-dependent function, $\lambda(\mathbf{s})$. This, such that the expected number of points, $E(N(\Omega))$, of the Poisson process in some bounded region, Ω, is given by $\int_\Omega \lambda(\mathbf{s})\mathrm{d}s < \infty$. Herein, and in the remaining chapters this intensity measure $E(N(\Omega))$, also denoted $\Lambda(N(\Omega))$, is considered to be random. This leads to a doubly-stochastic Poisson process, also called a Cox process due to its introduction by (Cox and Isham, 1980). A special case of this process the LGCP—introduced in Section 1.2.2—requires that the logarithm of the intensity measure is a Gaussian process. That is, the intensity measure is a realisation of a positive random measure. Leading to the number of points within Ω, $N(\Omega)$ being given by,

$$N(\Omega) \sim \text{Poisson}\left(\int_\Omega \lambda(\mathbf{s})\mathrm{d}s\right),$$

where $\lambda(\mathbf{s})$ is random and the logarithm of a Gaussian process, $z(\mathbf{s})$. Simple LGCPs may be fitted through the use of minimum contrast approaches (Møller and Waagepetersen, 2007). However, to fit LGCP with more complex dependency structures traditionally computational prohibitive MCMC methods were used. This, mainly prior to methodology based on Laplace approximations being introduced by Rue et al. (2009), (Taylor and Diggle, 4). The use of this approach in fitting LGCPs is considered in this and the following chapters, utilising techniques allowing more flexibility in the intensity measure.

The Gaussian process, $z(\mathbf{s})$, in 2 dimensions, is termed a GRF, or a GMRF in the discrete

case. A GRF, is simply a collection of multivariate random variables defined with respect to their neighbourhood structure. Historically, mainly due to computational burden, point patterns were binned into grids. Doing so results in a loss of spatial information. Yet, through this discretisation the intensity measure—the log of the Gaussian process—can be fitted with computational feasibility. Yet, recent methodology introduced by Lindgren et al. (2011) links GMRFs to a GRF, through the solution of a SPDE, (introduced in Section 1.3.2). This approach enables one to consider a continuously indexed Gaussian process (GRF).

The following sections detail fitting LGCP models, through the use of Integrated nested Laplace approximation methodology introduced by Rue et al. (2009). Moreover this methodology is coupled with the SPDE approach, (Lindgren et al., 2011), enabling a continuous intensity measure to be assumed for the latent GMRF. In particular, this chapter introduces functionality of the `R` package `lgcpSPDE`[1]. The `lgcpSPDE` package is based on the fitting methodology—Laplace approximation—of `R-INLA`. However, it focuses on the fitting of LGCP models, specifically, concentrating on the importance and interpretation of the intensity measure. The latent GRF that is assumed in a LGCP—the exponential of the stochastic intensity—encodes far more information regarding the structure of the point pattern, than simply being a reflection of the observed point arrangement. To maintain notational consistency with Section 1.1, the mean (Poisson) number of points is denoted $\Lambda(\mathbf{s})$. This spatially varying function can be thought of as an amalgamation of all processes operating to proffer the structure of the observed point pattern.

This chapter chiefly focuses on estimating $z(\mathbf{s})$, thus $\Lambda(\mathbf{s})$ (its role in the fitting of LGCPs elucidated further in Section 4.1.1). Section 4.2 introduces the use of `R-INLA` (Rue and Martino, 2007), in order that the model fitting stages taken in the following sections and the applications in the following chapters are established. Section 4.3 details the functionality of the `R` package developed by the author in the fitting of LGCP models.

4.1.1 The intensity measure of a log-Gaussian Cox process

Recall that the intensity process, $\Lambda(\mathbf{s})$, is given by,

$$\Lambda(\mathbf{s}) = \exp(z(\mathbf{s})) = \int_{\Omega} \lambda(\mathbf{s}) \mathrm{d}s,$$

where $z = (z(\mathbf{s}) : s \in \mathbb{R}^2)$ is a real-valued Gaussian process, that is, the joint distribution of any $(z(s_1), \ldots, z(s_n))$ is Gaussian. Thus,

$$z(\mathbf{s}) \sim N(\boldsymbol{\mu}, \mathbf{Q}^{-1}),$$

as given in Equation (1.4). Specifically $z(\mathbf{s})$ is normally distributed with mean $\boldsymbol{\mu}$ and precision $\mathbf{Q}$. In addition, $z(\mathbf{s})$ is assumed to have a Matérn covariance, given in Equation (1.6). Of

[1] which is not yet available on CRAN but available at `https://github.com/cmjt/lgcpSPDE`

course, any positive definite covariance function may be considered—a constraint of defining GRFs (Cressie, 1993)—however a consequence of using the SPDE methodology to model $z(\mathbf{s})$ requires that the covariance function is Matérn, see (Lindgren et al., 2011).

Commonly, natural processes over which inferences are sought are not constant in time, as such the latent structure may be spatio-temporal in nature. Specifically, whereby $\mathbf{z}(\mathbf{s}, t)$ follows a first order auto-regressive structure in time with coefficient ρ and is given by;

$$\mathbf{z}(\mathbf{s}, t) = \rho\, \mathbf{z}(\mathbf{s}, t-1) + w(\mathbf{s}, t), \tag{4.1}$$

where $w(\mathbf{s}, t)$ is spatio-temporal white noise, for $i = 1...n$, $t = 1, ..., T$. Such a model is a separable space-time model, and is defined by the Kronecker product between the precision of a SPDE model for the spatial domain and an AR(1) for the temporal domain. Leading on from the introduction to fitting LGCP models detailed in this chapter, this thesis goes on to detail the development of flexible marked spatio-temporal point process models. Applications are considered in a wide range of fields, where the similarities in the data structure enable models of the same mathematical construction to be fitted, resulting in inferring complex dependencies inherent in many marked point pattern datasets

The following sections are directed at estimating $z(\mathbf{s})$, thus, inferring $\Lambda(\mathbf{s})$. Specifically, introducing functionality of the `R` package `lgcpSPDE` in fitting LGCPs and visualising $\Lambda(\mathbf{s})$. To illustrate the model fitting procedure and inference, data pertaining to the recorded locations of terrorist events obtained from an open-source online database Global Terrorism Database (GTD), and supplied with the package `lgcpSPDE`[2] as the `R` object `terrorism`, are used throughout Section 4.3. Work relating to analysis of this data using similar methodology to that discussed herein has recently been submitted to the *Journal of the Royal Statistical Society: Series A (Statistics in Society)* (*Python, A., Illian, J.,* ***Jones-Todd, C.****, & Blangiardo, M.* "A Bayesian Approach to Modelling Fine-Scale Spatial Dynamics of Non-State Terrorism: World study 2002-2013"). This manuscript primarily focuses on quantifying the spatial dynamics of terrorism and its uncertainty across the world. Specifically, utilising a latent GRF in modelling the spatial dynamics inherent in the probability of a terrorism event being fatal. The contribution of this thesis' author to this manuscript was in the development of the code required to fit the spatio-temporal model, akin to functionality of the package `lgcpSPDE`. However, as this chapter is focused on the fitting of LGCPs the same data is used to illustrate the model fitting procedure and the essence of the estimated GRF.

4.2 Using `R-INLA` and the SPDE approach in fitting a LGCP

This section introduces the use of the package `R-INLA` (Rue and Martino, 2007) in fitting a basic predictive LGCP model, with a single fixed covariate. This section is purely for

[2]available at `https://github.com/cmjt/lgcpSPDE`

illustration, no results are discussed, rather the functionality offered by `R-INLA` in fitting and obtaining output from such models. The approach is demonstrated using point pattern data (simulated in the unit square), the aim being to illustrate on a step by step basis the requirements for fitting a latent Gaussian point process model—a LGCP—using the INLA-SPDE approach.

Representations in `R-INLA`

Within `R-INLA` the parameters of the SPDE model detailed in Section 1.3.2 are represented by $\log(\tau)= \theta_1$, and $\log(\kappa)= \theta_2$, where θ_1 and θ_2 are assigned a joint normal prior distribution. However, typically it is more logical to talk about the standard deviation, σ, and range, η, of the random field. The range, $\eta = (8\nu)^{1/2}/\kappa$, of a field is the distance at which the correlation falls to $\sim 0.13, \forall \nu > 1/2$. Thus, the link to τ, and κ in `R-INLA` is;

$$\begin{aligned} \log \tau &= \frac{1}{2}\log\left(\frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}}\right) - \log\sigma - \nu\log\kappa, \\ \log \kappa &= \frac{\log(8\nu)}{2} - \log\eta. \end{aligned} \tag{4.2}$$

Reparmeterising the representations in Equation 4.2 by using baseline standard deviation, σ_0 and range, η_0, values gives `R-INLA`s internal representations;

$$\begin{aligned} \log \kappa_0 &= \frac{\log(8\nu)}{2} - \log\eta_0, \\ \log \tau_0 &= \frac{1}{2}\log\left(\frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}}\right) - \log\sigma_0 - \nu\log\kappa_0, \\ \log \tau &= \log\tau_0 - \theta_1 + \nu\,\theta_2, \\ \log \kappa &= \log\kappa_0 - \theta_2, \end{aligned}$$

hence θ_1 and θ_2 jointly control τ.

The use of a SPDE as an approximation to a GMRF was introduced by Lindgren et al. (2011). Here, weighted sums of basis functions—constructed using the Delauney triangulation of the region— are used to approximate the spatial random functions arising from the solution to the considered SPDE discussed by Lindgren et al. (2011). Thus, the continuous interpretation of space is preserved, whilst benefiting from the computational advantages arising from the discrete Markovian structures of GMRFs. In this setting, the locations are projected onto the mesh, discussed in Section 1.3.2 (see Figure 1.3). The response is therefore a vector of length equal to the number of observations and the number of mesh nodes, where the response equals 1 at the observation locations, and 0 at the mesh nodes. The finite basis representation, given by Equation 1.7, and illustrated by Figure 1.3, approximates the value of the GRF which is assumed to be the value of the intensity surface at that particular location.

Section 4.2.1 below leads the reader through a step by step guide to fit a predictive, single covariate LGCP model. In addition providing code which may be lifted, and used for any data with the same structure.

4.2.1 Model fitting procedure

Initially R Example 4.2.1 simulates a homogeneous Poisson process in the unit square—note, this purely demonstrates the structure of the required data in fitting the model described below—as well as a spatial polygon describing the boundary of the observed point pattern (these are the `R` objects `locs` and `domain` respectively). A spatial polygon of the domain is not required to fit a LGCP in this context, however, typically in certain fields (e.g., ecology, geology), exact domain (survey region) is known, thus, incorporation of this structure is illustrated herein.

In order to fit a LGCP model using the INLA-SPDE approach all that is required are the point locations (either in 2-dimensions, or defined on a sphere, which may represent the globe). The point pattern discussed here is simulated within the unit square, however, any geo-referenced data may be used (that is, any recognised coordinate system). Therefore, to fit a LGCP model (in 2-dimensions), the object `locs` as in R Example 4.2.1 should be a $2 \times \text{n}$ matrix (where n is the number of observations). The first six rows of this object are shown, and used in each demonstrative step below. As mentioned above, a spatial polygon—object `domain` in R Example 4.2.1—is not required, but, may be used in the modelling procedure if available. This section describes in detail how this information is incorporated; whilst also demonstrating how it is not a necessity for the modelling procedure.

R Example 4.2.1.

```
locs <- cbind(runif(300,0,1),runif(300,0.1))
domain <-SpatialPolygons(list(Polygons(
    list(Polygon(cbind(c(0,0,1,1),c(0,1,1,0)))),"s1")))
class(domain)
## [1] "SpatialPolygons"
## attr(,"package")
## [1] "sp"
head(locs)
##            [,1]      [,2]
## [1,] 0.09878282 0.2997633
## [2,] 0.48823179 0.9759309
## [3,] 0.36403673 0.5047489
## [4,] 0.42061913 0.6993227
## [5,] 0.30096439 0.2532919
## [6,] 0.14763513 0.7394571
```

Construction of the `mesh`

Construction of a Delauney triangulation (mesh) is the first step to fitting a LGCP in the INLA-SPDE context. The construction of the triangulation is an influential step in model fitting (Lindgren, 2012), as discussed in Chapter 1. The spatial resolution of the triangulation affects inference, In particular construction of the mesh should reflect the correlation range of the assumed latent process. That is, if the spatial structure is unchanging then the triangles need not be fine due to the values of the process at neighbouring locations likely to vary little from those observed. However, if there is a lot of local scale variation, the triangles must be small in order to capture this variation.

Demonstrated in the section headed `CREATE MESH` in R Example 4.2.2, the mesh construction is done using the `inla.mesh.2d()` function. The bottom plot in Figure 4.2 illustrates the triangulation constructed through running the code in R Example 4.2.2. Here, the spatial polygon of the domain, supplied to the `boundary` argument to `inla.mesh.2d()`, is used to create the `R` object `mesh` (used in all sections entitled R Example). Moreover, the top row of Figure 4.2, illustrates two further constructed mesh, both created using the point pattern locations (supplied to the `loc` argument to `inla.mesh.2d()`), as opposed to the spatial polygon of the domain (either `loc` or `boundary` must be supplied).

Further arguments of the function `inla.mesh.2d()` control the size of the triangles within the mesh. The two main arguments used to control the size of the triangles are(i) `cutoff`: (optional) which refers to the length at which to cut off triangle edge lengths, (ii) `max.edge`:

(mandatory) a vector of length two specifying the triangle edge lengths inside and outside the region respectively (if a single numeric value is supplied this refers to the area inside the region itself). The magnitude of the numeric values supplied to these arguments, clearly must reflect the order of the spatial coordinates supplied (e.g., a triangle edge cutoff of 500, is pointless if the domain is the unit square). The triangulations illustrated in the top row of Figure 4.2 demonstrate how changing the values of these arguments affect the triangle sizes. The code `inla.mesh.2d(loc = locs, max.edge = c(.3,0.6), cutoff = .3 )`, and `inla.mesh.2d(loc = locs, max.edge = c(.09,.1),cutoff=.4)` is used to construct the left and right mesh respectively.

How choice of triangulation may affect estimator performance

As discussed above the choice of triangulation can drastically affect estimator performance. This as the triangulation represents a discretisation of the domain forming the deterministic basis functions, $\phi_k(\mathbf{s})$ given in Equation 1.7. To illustrate how the scale and shape of the triangulation affects parameter performance, six triangulations were constructed over the domain of the simulated data discussed in this section, and the percentage difference from the estimated and true parameters calculated (Figure 4.1). It is important to note that typically the triangulation is constructed outwith the boundary in order to prevent boundary effects. However, in the case of a point process model this decreases estimator's performance, see the meshes with both and inner and outer edge in Figure 4.1. This, as the SPDE model for the GRF is defined on the nodes of the mesh which—in the fitting of a LGCP— are considered the integration points. Thus, essentially the area of the triangle is considered to be proportional to the expected number of events. Thus, inclusion of nodes outwith the domain results in a larger expected value. Hence, when fitting a LGCP it is advisable to only construct the mesh over the domain (as in the other four meshes in Figure 4.1). The other properties of the mesh affecting the estimator performance are the spatial scale of the triangles, which clearly affects model fitting time too. Generally, a finer resolution of mesh will improve estimation but increase computational time. However, dependent on the smoothness of the process giving rise to the observed pattern the triangulation need not be unduly fine. As there exists no precise measure of an appropriate triangulation—a trade off between estimation accuracy and computational time—it is highly recommended that various meshes are trialed reflecting the spatial scale and smoothness of the assumed process.

Build the `projector matrix`

The triangulation—consider the `mesh` object in R Example 4.2.2—is a (basis) representation of the domain (see Section 1.3.2). Once the `mesh` has been constructed, a projector matrix must be constructed. This links the observations (`locs`), and the `mesh`, giving the neighbourhood structure of the observations (`locs`). This is demonstrated in the section headed `BUILD`

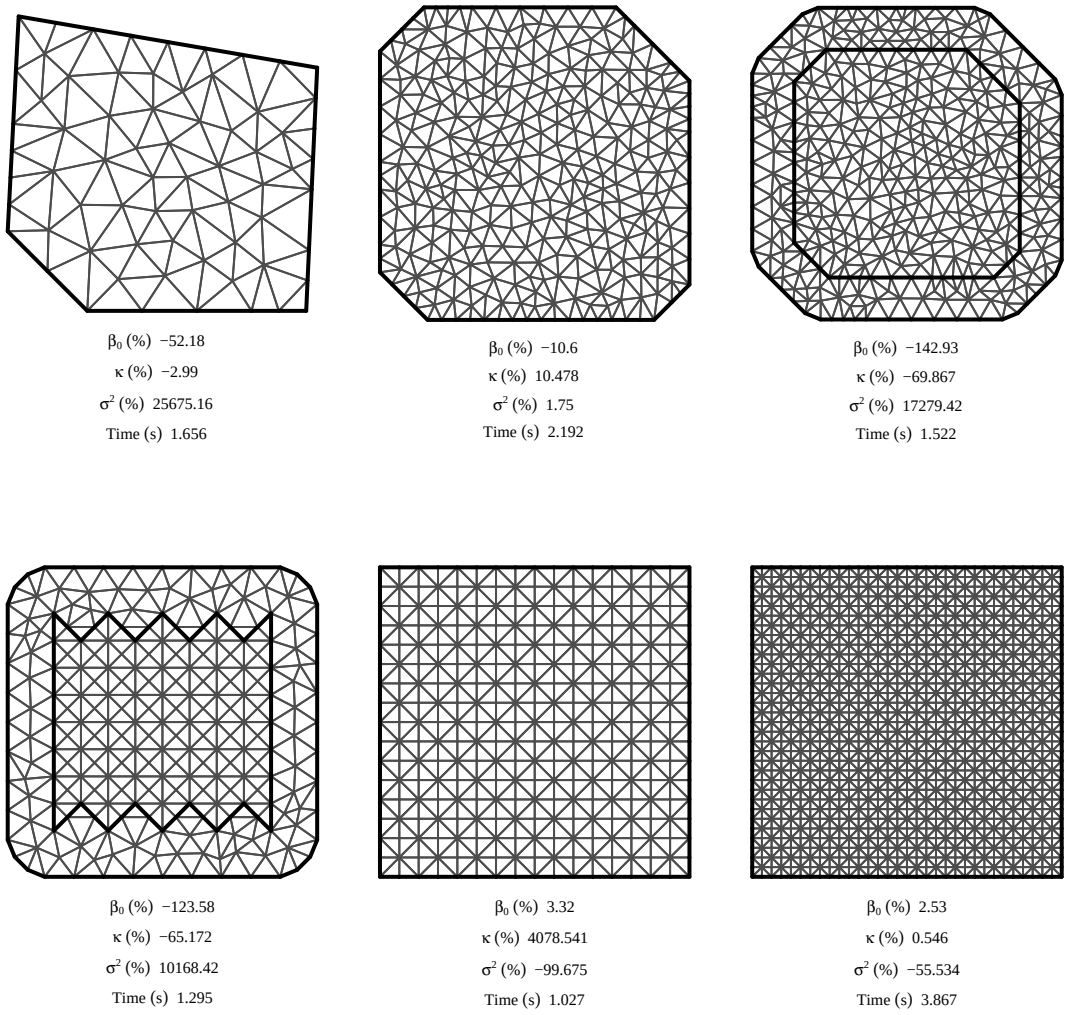


Figure 4.1: Plots of each Delauney triangulation (mesh) used to compare estimator performance fitting a LGCP model to the simulated data `lgcp2D` using the INLA-SPDE approach. Below each mesh the percentage difference from the true parameter values used to simulate the point pattern as well as the total time used (in seconds) to fit the model (on a laptop with Intel Core i7 processor).

`PROJECTOR MATRICES` in R Example 4.2.2, using the `inla.spde.make.A()` function to create the projector matrix object `Ast`. The two arguments of this function are unsurprisingly the two objects—the triangulation (`mesh`) and the observations (`locs`)—one wishes to link.

It is worth noting that when fitting any point process model (a LGCP is being fitted in this case) the observations are the point locations themselves. Yet, if say a geostatistical model were being fitted, the observations being some continuous variable, the argument `loc` in the R Examples given here would refer to the observation locations (which in the case of point pattern data are one and the same).

In addition to the projector matrix `Ast`, another projector matrix `Ast.pred` is created in the code provided in R Example 4.2.2, this is such that prediction can be carried out during

the estimation process. The locations for prediction, are (purely for purposes of example) chosen to be the first ten locations of the observations (`locs[1:10,]`), but could be any locations within the domain. R Example 4.2.3 and the subsections below, go on to explain how this is used in the prediction.

Define the `spde` model

The SPDE model is defined on the basis representation (`mesh`), and is defined through the function `inla.spde2matern()`, which takes the triangulation (`mesh`) as its only argument (in this setting). This is illustrated in the section headed `DEFINE SPDE MODEL` in R Example 4.2.2.

Build the `data stack`

R Example 4.2.3, headed `DEFINE VARIABLES AND BUILD DATA STACK` is perhaps the most involved in the modelling fitting procedure discussed herein. The object `y.pp` is defined as the response variable, it is set to be 0 at the `nv` mesh nodes, and 1 at the `n` observation locations. Thus, the SPDE model for the intensity surface essentially interpolates around these locations giving an estimate of the intensity surface over the domain.

For illustrative purposes only the `covariate` vector is simulated, this is to demonstrate that the covariate vector must be of length (`nv` + `n`), that is collected at both the `mesh` node locations, `mesh$loc`, and the observation locations, `locs`.

The object `expected` is specific for fitting a point process model, and can be thought of as an effort variable which is obtained through information contained in the basis representation, on which the SPDE model is constructed, `spde`. The spatial effect—`field.pp`—is defined on the basis representation (triangulation), and is simply an index from 1 to the number of `mesh` nodes, `nv`, which is then linked to the neighbourhood structure by the projector matrix, `Ast`.

Finally R Example 4.2.3 demonstrates constructing the data stacks for both the observations, `stk.pp`, and the predictions, `stk.prd`. Then, combining both into one data stack, `stack`. Supplied to the function `inla.stack()` are,

- the `data`, supplied as a list, for the observations both the response vector, `y.pp`, and 'effort' vector, `expected`, are given. Whereas for the prediction a list of `NA`s are given, as during the estimation procedure `R-INLA` will predict at the prediction locations defined in the projector matrix `Ast.pred`.

- A list of projector matrices, supplied through the argument `A` where each element of the list is the corresponding link—projector matrix—between the locations of the response (or predictions), in the case of the actual observations this is `rBind(Diagonal(n=nv), Ast)`. Here `Diagonal(n=nv)` is the link between the `nv mesh` nodes and the `mesh` (as

the first `nv` observations of the response in the point process setting are 0's at the mesh nodes). In the case of prediction, the link (between the predication locations and the triangulation (`mesh`)) is defined through the projector matrix `Ast.pred`,

- the `effects`, supplied as a list containing both random (the spatial random effect `field.pp`) and a list of fixed effects (`intercept` (simply a vector of 1's) and `covariate` (both of length `nv + n`)). In the case of prediction there is only one random effect (again defined at the `mesh` nodes),
- a `tag`, (optional) useful in order that post model fitting, information regarding locations of interest can be extracted, in order to carry out inference.

Define `formula`

The syntax that can be seen in R Example 4.2.4 under the header `DEFINE FORMULA` is used to define the model in `R` as understood by `R-INLA`. The fixed effects (`intercept` and `covariate`) are defined in the data stack, `stack` in R Example 4.2.3. Here, the wrapper `f()`, defines some random effect, which in this case is named `field.pp` (as defined in the data stack R Example 4.2.3). The spatial random effect is defined using `f(field.pp,model = spde)`, and is modelled using a SPDE model defined as the object `spde`, R Example 4.2.2.

Call to `inla()`

The final step is the call to `inla`, and is illustrated in the section headed `CALL TO INLA` in R Example 4.2.4. Mandatory arguments (in this setting) are a `formula` object,`data = inla.stack.data(stack)`, and `control.predictor = list(A = inla.stack.A(stack), compute = TRUE)`. The latter two arguments, inform the model fitting procedure as to the data structure (which is contained in the object `stack`). The `family` argument refers to the likelihood of the observations, by default this is `''gaussian''`. However, in the point process context this is `''poisson''` as in this context when modelling a point pattern.

The argument `E=inla.stack.data(stack)$e` is required when fitting a point process model. This is akin to the effort, required for a Poisson likelihood. Finally the argument `verbose` (by default `FALSE`) is Boolean, prints out the model fitting iterations to the console.

R Example 4.2.2.

```
### CREATE MESH
bound <- inla.sp2segment(domain)
mesh <- inla.mesh.2d(boundary = bound, max.edge = c(.1,.12),cutoff=.12)
### BUILD PROJECTOR MATRICIES
## create projection matrix for observation locations
Ast <- inla.spde.make.A(mesh = mesh, loc = locs)
## create projection matrix for prediction locations
Ast.pred <- inla.spde.make.A(mesh = mesh, loc = locs[1:10,])
### DEFINE SPDE MODEL
spde <-inla.spde2.matern(mesh = mesh)
```

R Example 4.2.3.

```
### DEFINE VARIABLES AND BUILD DATA STACK
# number of observations
n <- nrow(locs)
# number of mesh nodes
nv <- mesh$n
## the response for the point pattern locations
y.pp <- rep(0:1, c( nv, n))
## simulate covariate (which must be defined at both the mesh node
## locations and the observations)
covariate <- cos(c(mesh$loc[,1],locs[,1])) +
    rnorm(length(c(mesh$loc[,1],locs[,1])))
## effect for LGCP used for point pattern
st.volume <- diag(spde$param.inla$M0)
expected <- c(st.volume, rep(0, n))
## create the spatial random effect index
field.pp  <- 1:nv
## data stack for the observations
stk.pp <- inla.stack(data=list(y=y.pp, e=expected),
                     A=list(rBind(Diagonal(n=nv), Ast),1),
                     effects=list(field.pp = field.pp,
                                  list(intercept = rep(1,nv+n),
                                       covariate = covariate)),
                     tag = 'observation')
## then for the prediction locations
stk.prd <- inla.stack(data=list(y=NA), A=list(Ast.pred),
                      effects=list(field = 1:spde$n.spde),
                      tag='prediction')
## combine data stacks
stack<-inla.stack(stk.pp,stk.prd)
```

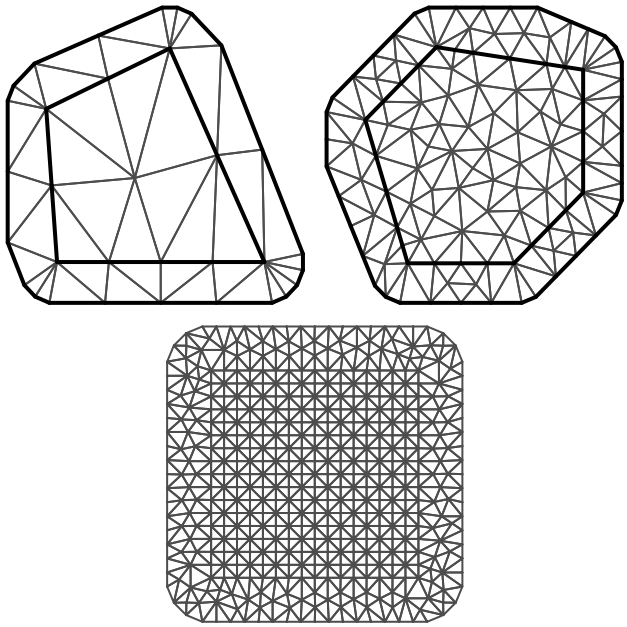


Figure 4.2: Three Delauney triangulations over some domain, where either the boundary information is supplied (Top row) or not (bottom plot). Top row: left hand mesh is constructed through running the following code `inla.mesh.2d(loc = locs, max.edge = c(.4,0.7), cutoff = .5 )`, right hand mesh constructed through running the following code `inla.mesh.2d(loc = locs, max.edge = c(.15,.2),cutoff=.4)`. The triangulation on shown on the bottom row is constructed using the code in R Example 4.2.2, creating the `R` object `mesh` which is used in the following R examples in this section.

R Example 4.2.4.

```
## DEFINE FORMULA
formula <- y ~ 0 + intercept + covariate + f(field.pp, model=spde)
## CALL TO INLA
 ##call to inla
result <- inla(formula, family = "poisson",
               data=inla.stack.data(stack),
               E=inla.stack.data(stack)$e,
               control.predictor=list(A=inla.stack.A(stack),
                                      compute = TRUE),
               verbose = TRUE)
```

4.2.2 Inference

The object `result` returns a list comprising of the fitted model components, and model fitting procedure. Calling `summary()` on the fitted model will return a summary of both the fixed and random effects. In the instance of the object `result` above the fixed effects would be the intercept (`intercept`) term and the coefficient to `covariate`. The parameters of the random

effects (here `field.pp`) are also returned. Densities of the parameters are plotted through calling `plot()` on the object. However, typically when fitting latent Gaussian models the interest lies in the structure of the spatial field (here `field.pp`). R Example 4.2.5 illustrates how to extract the estimated mean values of this random field, which then can be plotted using the command `image.plot(proj$x, proj$y, rf_mean)` (in the package `fields`). To visualise the uncertainty of the field the standard deviation can also be extracted through `inla.mesh.project(proj, field = result$summary.random$field.pp$sd)`. Moreover, the latter part of R Example 4.2.5 demonstrates how to only plot the values of the random field inside the domain (recall that the `mesh` is extended beyond the boundary to prevent boundary effects).

R Example 4.2.5.

```
## extracting the estimated mean of the random field at the mesh verticies
proj <- inla.mesh.projector(mesh)
rf_mean <- inla.mesh.project(proj,
                             field = result$summary.random$field.pp$mean)

## the following only applies when boundary information is available
## in the form of a spatial polygon
library(spatstat)
library(maptools)
in.domain <- function(proj, domain){
    e<-expand.grid(proj$x,proj$y)
    w<-as.owin(domain)
    o<-inside.owin(e[,1],e[,2],w)
    o<-matrix(o,nrow=length(proj$x))
    return(o)
    }
inside <- in.domain(proj = proj, domain = domain)
rf_mean[!inside] <- NA
```

Following on from detailing existing methodology in the fitting of LGCP models using the INLA-SPDE approach Section 4.3 introduces the readers of this thesis to the functionality of the `lgcpSPDE` package developed by the author of this thesis. The objective of the above section was to give the reader a basic understanding of the use of the model fitting procedure upon which the functionality of `lgcpSPDE` is based.

4.3 The lgcpSPDE package

The lgcpSPDE package developed by the author of this thesis offers functionality based on INLA methodology in the fitting of LGCP models. In addition, further methodology is developed for the fitting of joint marked point process models. Such methodology, and extensions are discussed in the chapters which follow on from this.

The R package lgcpSPDE can be installed following instructions available from https://github.com/cmjt/lgcpSPDE[3]. The lgcpSPDE package contains an example dataset, pertaining to the GTD data (the object terrorism). This dataset is analysed in the recently submitted article Python et al. (2016). In addition, this data is used throughout this section to demonstrate use of the package in fitting both spatial and spatio-temporal LGCPs and visualising the spatial structures relating to the intensity measure of the process.

4.3.1 Model fitting

The make.mesh() function

Fitting an LGCP using INLA and an SPDE model for the latent field requires triangulation of the domain, as discussed in Section 1.3.2. The make.mesh() function, provided by lgcpSPDE, is essentially a wrapper function for functions provided by INLA, and has the following arguments:

- locs: A $n \times d$ matrix of n locations in $d = 1, 2, 3$ dimensions. Either this or the argument spatial.polygon must be supplied.
- mesh.pars: a named vector of mesh parameters, must contain cutoff length at which to cut off triangle edge lengths, min triangle edge length inside region, and max triangle edge length inside region.
- spatial.polygon: If supplied the spatial polygon for the domain is used to construct mesh,
- sphere: Logical if TRUE the mesh is constructed on the unit sphere, note this is only possible if coordinates are Longitude and Latitude. By default this is FALSE.

R Example 4.3.1.

```
data(terrorism)
locs <- terrorism[,1:2]
mesh.1 <- make.mesh(locs = locs,
                    mesh.pars = c(max = 50, min = 150,
```

[3]This package is still under development, therefore documentation is relatively sparse

```
                                  cutoff = 50 ))
mesh.2 <- make.mesh(locs = locs,
                    mesh.pars = c(max = 0.3, min = 0.1,
                                  cutoff = 0.4 ),sphere = TRUE)
mesh.3 <- make.mesh(spatial.polygon = studyarea,
                    mesh.pars = c(max = 20/180,
                                  min = 5/180, cutoff = 5/180 ),
                    sphere = TRUE)
mesh.4 <- make.mesh(spatial.polygon = studyarea,
                    mesh.pars = c(min =  10/180, cutoff =  10/180),
                    sphere = TRUE)
```

The `fit.lgcp()` function

The `fit.lgcp()` function facilitates the fitting of spatial or spatio-temporal LGCPs using the INLA-SPDE approach. The main arguments of `fit.lgcp()` are:

- `mesh`: a mesh object (i.e., Delauney triangulation of the domain), an object returned by `make.mesh`. It is highly recommended that this is supplied, as choice of triangulation can drastically affect estimator performance (see below). If this is not supplied the function does have inbuilt capacity to create a heuristic triangulation based on the argument `mesh.pars` or is this not supplied on the point locations alone (definitely not recommended, however is an option for demonstrative reasons).
- `mesh.pars`: a named vector of mesh parameters, must contain `cutoff` length at which to cut off triangle edge lengths, `min` triangle edge length inside region, and (optionally) `max` triangle edge length outside region. If no triangulation—`mesh`—is supplied, then these parameters are used (calling the function `make.mesh()`) to construct the triangulation.
- `locs`: required, a matrix of observation locations, where each row corresponds to the observation. If neither `mesh` nor `mesh.pars` is supplied then locations are used to construct a (unrefined) triangulation.
- `temp`: a numeric vector specifying a temporal index for each observation $(1, \cdots, T)$
- `covariates`: a named `R data.frame` of covariates, where for each covariate vector the first m must be the covariate values at the m mesh nodes(if not available can be supplied as `NA`s), and the latter n be the values at the n observation locations. Thus, if covariates are included in the analysis the user must also supply a constructed mesh in order that the mesh nodes can be extracted (`mesh$loc`).

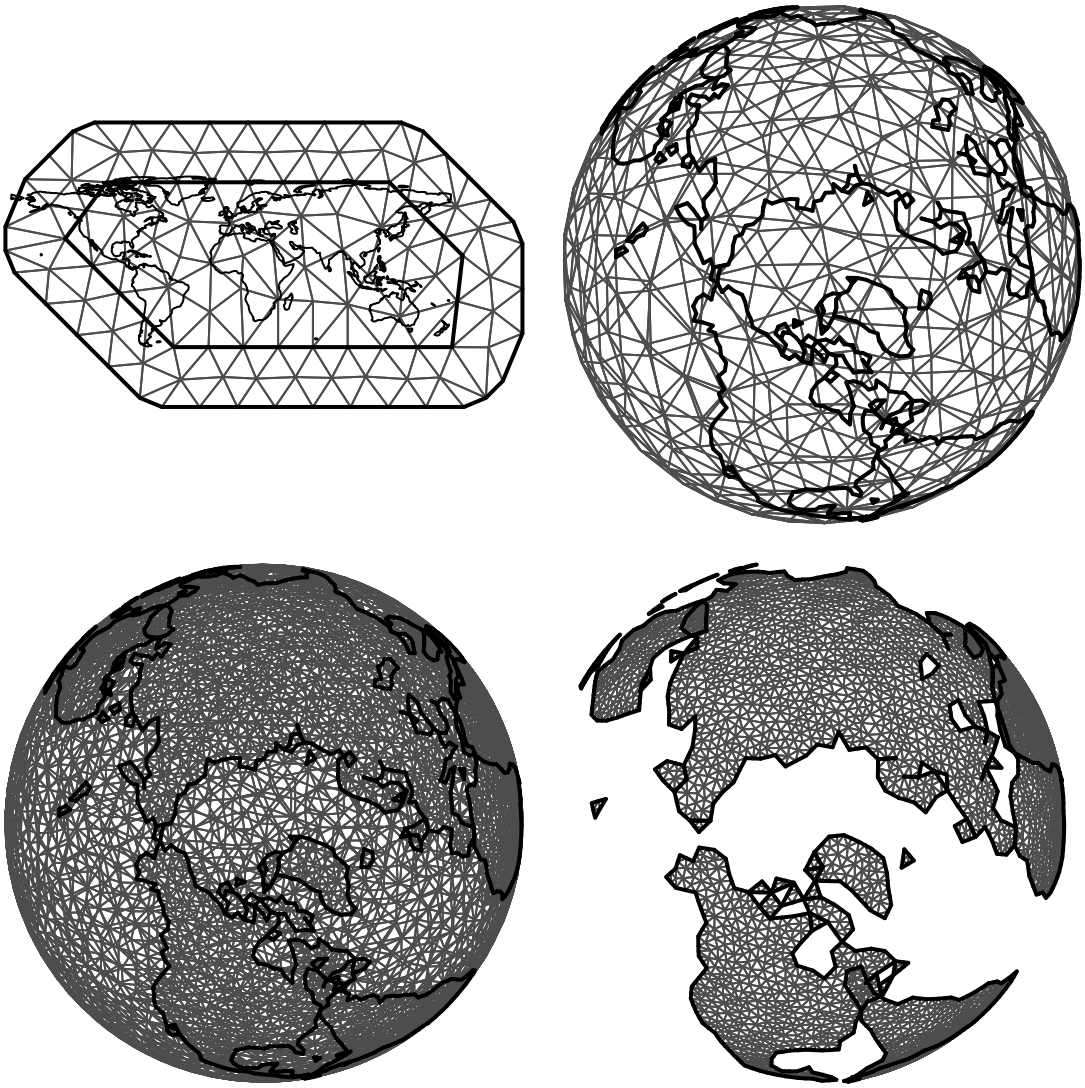


Figure 4.3: Plots showing four different mesh constructions—R Example 4.3.1—for the `terrorism` dataset. Top left: The Delauney triangulation—`mesh.1`—assuming a 2D domain, built using the Latitude and Longitude of the recorded `terrorism` data. In all the remaining plots the coordinates are projected onto the unit square, upon which the triangulation is constructed. Top right: `mesh.2`, constructed using the coordinates of events (projected onto the unit sphere). Both the top row triangulations are overlaid with a spatial polygon of the world's continents, not used in the mesh construction. Bottom left: Triangulation constructed using the spatial polygon of the world's continents, note how the triangulation—`mesh.3`—over the land mass is made to be finer than that over the sea. Bottom right: Triangulation—`mesh.4`—solely over the land mass of the globe. The triangulations in the bottom row differ in one regard only, that being the construction or lack thereof over the seas. One may think that due to terrorism events (in this case) being purely land based, that the construction of the triangulation over the seas is pointless, and may affect inference. Yet, in the case where the triangulation solely exists over the land mass, inference will be affected by edge effects, that is greater uncertainly at the boundaries, which is circumvented when the triangulation is extended beyond the boundary.

Other arguments of this function are `prior.rho` which is the prior for the temporal correlation coefficient. By default a `pcprior` is used with `param=c(0-0.9)`, (`prior.rho = = list(theta = list(prior = "pccor1", param = c(0, 0.9)))`). This is a penalised complexity prior derived by Martins et al. (2014), which aims to penalise the complexity induced by deviating from the simpler base model. Another default argument is `verbose = FALSE`—logical, if `TRUE` model fit is output to screen. Moreover, the argument `return.attributes` is logical where if `TRUE` the Delauney triangulation—mesh—used in the model fitting procedure is returned as an additional attribute.

Running the following code—R Example 4.3.2—will fit a spatio-temporal LGCP to the terrorism dataset, where; `mesh` is the Delauney triangulation of choice(in this case `mesh <- mesh.4` from R Example 4.3.1), `locs` the Longitude & Latitude locations projected onto the unit sphere (as the Delauney triangulation is), `temp` a numeric vector of year indices—
`tmp <- as.Date(terrorism$Date)`
`temp <- as.numeric(format(tmp,'%Y')) -(min(as.numeric(format(tmp,'%Y')))-1)`.

R Example 4.3.2.

```
fit <- fit.lgcp(mesh = mesh, locs = locs, temp = temp,verbose = TRUE,
                return.attributes = TRUE)
fields <- find.fields(fit,n.t = 12,spatial.polygon = studyarea)
```

The `find.fields()` function

The `find.fields()` function extracts properties—mean or standard deviation—of the random field/s from an object obtained from calling `fit.lgcp()`. This function will also work on a model fitted in the conventional manner through a call to `inla()`. A nested list object is returned, where the first level of list pertains to each field, and the nested levels respectively pertain to each realisation of that field for example when a spatio-temporal model is fitted. Figure 4.4 shows the estimated mean of the GRF in 2002, resulting from the model fitted object `fit` in R Example 4.3.2. The main arguments of `find.fields()` are,

- `x`: a model fit object return by `fit.lgcp()`
- `mesh`: a mesh object i.e. Delauney triangulation of the domain used in the model fitting procedure, if `fit.lgcp()` is run with argument `return.attributes = TRUE` there is no need for this to be supplied
- `n.t`: a numeric vector of time indices, only supplied if `x` pertains to a spatio-temporal model fit
- `sd`: logical, if `TRUE` standard deviation of the GRF is returned as opposed to the mean

Figure 4.4: Estimated mean of the GRF in 2002 from the model `fit` in R Example 4.3.2. The spatio-temporal intensity measure for the model object `fit`, is in this case assumed to be given by $\Lambda(\mathbf{s},t) = \exp(z(\mathbf{s},\mathbf{t})) = \exp(\beta_0 + x(\mathbf{s},\mathbf{t}))$, where here $x(\mathbf{s},2002)$ is shown. Therefore to $\exp(\beta_0 + x(\mathbf{s},\mathbf{t}))$ (for each t $= 2002,\cdots,2012$) would show the spatially varying intensity function. Moreover, the temporal correlation between each year is estimated to be 0.993 (obtained by using the utility function `summary()` on the model object) indicating that there is strong spatial correlation over time in $x(\mathbf{s},t)$, thus, the intensity measure.

- `plot`: logical, if `TRUE` the required properties of the GRF are plotted (additional graphical parameters of the plot can also be supplied through)
- `spatial.polygon`: optional, if a spatial polygon of the domain is supplied, only values of the random field within the domain will be returned

In addition to the functionality discussed here, utility functions such as `summary()` will operate as on an object of class `inla`.

4.3.2 Simulating a log-Gaussian Cox process-the `rlgcpspde()` function

The `lgcpSPDE` package contains two sets of simulated data, which can be found in the exported objects `terrorism` (mentioned above) and `lgcp2D` . The latter provides a realisation of a LGCP simulated using a SPDE model, generated by the function `rlgcpspde()`[4]. The

[4]running `mu <- 1; kappa <- 0.5; sigma2 <- 0.05; spatial.polygon <- SpatialPolygons(list(Polygons(list(Polygon(cbind(c(0,0,10,10),c(0,10,10,0)))),"s1"))); mesh.pars <- c(min = 0.5, cutoff = 0.5); rlgcpspde(spatial.polygon = spatial.polygon,mu =`

function `rlgcpspde()`, returns a named matrix (or a list of matrices if spatio-temporal) of point locations and (if a marked point pattern is simulated) mark values. Its main arguments are as follows,

- `spatial.polygon`: the spatial polygon for the domain is used to construct the Delauney triangulation
- `mesh.pars`: a named vector of mesh parameters, must contain `cutoff` length at which to cut off triangle edge lengths, `min` triangle edge length inside region, and `max` triangle edge length inside region.
- `mu`: numeric, the intercept term to simulate a LGCP, by default is 0
- `kappa`: a numeric constant, parameter of the SPDE model which controls the spatial scale of the GRF
- `sigma2`: a numeric constant, parameter of the SPDE model, the marginal variance of the GRF, by default this is 0.05
- `n`: a numeric constant defining the number of time points, by default 1
- `rho`: the AR1 correlation coefficient for spatio-temporal samples, by default this is 0.9
- `mark`: Logical, if TRUE a marked point pattern is simulated
- `beta`: a scalar, this, the interaction parameter describing the dependence between the mark and point locations
- `mark.function`: a function of 2D spatial coordinates which describes the spatial process specific to the mark, by default this is `function(x,y) cos(x) - sin(y)`.
- `seed`: seed for the simulation, by default this is 1

As discussed in Section 4.1 in addition to the observed point pattern, there often exist attributes of the points (e.g., tree height), which are themselves realisation of some process, $M(\cdot)$. This mark process, $M(\cdot)$, is conditional on the process giving rise to the observed pattern. In addition to simulating a typical realisation of a LGCP, `rlgcpSPDE()` supports simulating—either a spatial or spatio-temporal—marked LGCP. R Example 4.3.3 illustrates simulating a marked LGCP with parameterisation $\Lambda(\mathbf{s}) = \exp(\beta_0 + x(\mathbf{s};\theta))$, where $\beta_0 = 1$ and $\theta = (\kappa, \sigma^2) = (0.5, 0.05)$—the same as those used in the simulated data object `lgcp2D`. Moreover, $\mathrm{M}(\mathbf{s}) = \mathrm{M}(x, y) = \cos(x) - \sin(y) + \beta\,\Lambda(x, y)$, where $\beta = 1$ is the interaction parameter describing the dependence between the mark and point locations. Other than to illustrate functionality of the `lgcpSPDE` package in simulating marked LGCPs, this chapter

`mu, sigma2 = sigma2, mesh.pars = mesh.pars, kappa = kappa,seed = 1)` will simulate the same pattern

does not refer further to marked point processes, such process models are discussed further in Chapters 5 and 6.

R Example 4.3.3.

```
locs <- rlgcpspde(spatial.polygon = spatial.polygon,mu = mu,
                  sigma2 = sigma2,mesh.pars = mesh.pars,
                  kappa = kappa,mark = TRUE,beta = 1)
head(locs)
##             x        y       mark
## [1,] 9.051680 8.218507 -0.6496963
## [2,] 5.826895 6.595438  1.3723254
## [3,] 4.318117 6.189785  1.9253244
## [4,] 4.151858 5.394370  1.4394891
## [5,] 3.956278 6.672369  1.6431902
## [6,] 6.212916 9.137036 -0.9242240
```

4.4 Discussion

The above chapter takes a slightly different tack to Chapters 2 and 3, predominantly in the methodology used to fit point process models. That is, the use of Integrated nested Laplace approximations in conjunction with Bayesian methodology over an approximate-likelihood approach. Despite the seeming differences between the structure of the processes discussed—NSPP, void process, and LGCP—all are hierarchical processes, where some latent process affects the observed structure of the point pattern. In the case of the LGCP this latent process—a GRF—reflects the logarithm of the intensity measure of the process. Therefore, the focus of this chapter has been on estimating the assumed latent Gaussian process $z(\mathbf{s})$ which reflects the intensity measure, $\Lambda(\mathbf{s})$, of the LGCP. The modelling procedure employed demonstrated the fitting of LGCP models using the INLA-SPDE approach, whilst demonstrating functionality of the `lgcpSPDE` package in doing so.

Using a SPDE model for the latent GRF requires that the assumed continuous process is approximated by some discretisation of the domain. This is achieved through a Delauney triangulation, which then provides the required basis functions $\phi_k(\mathbf{s})$ and weights w_k (Equation (1.7)). Specifically, where each mesh vertex has a weight, thus one triangle in the triangulation will have three and for a point within this triangle the projection is the weighted average. Moreover, the area of the triangle is considered to be proportional to the expected number of events. Thus, the value of the GRF at location i, $z_i(\cdot)$, is the product of the Gaussian weight and basis function at that location, which represents the logarithm of the intensity surface—in the case of a LGCP—illustrated by the pyramid in Figure 1.3. Thus, as noted

above the choice of triangulation is an integral step in the fitting of SPDE. This has been explored to some extent herein.

The remainder of this discussion summarises the steps required in the fitting of such models described above. Particular emphasis is put on the interpretation of the GRF. This structure in later chapters takes on a more influential role. Specifically, as methodology developed by the author of this thesis facilitates inferring additional processes operating conditional on the process proffering the observed structure of the point pattern. In particular, through inferring mark processes, $M(\mathbf{s})$.

4.4.1 Mechanisms operating in a LGCP

The intensity measure of a LGCP, given in Section 4.1.1, is simply a spatially varying function which returns the expected number of points within a bounded region. A LGCP assumes a latent Gaussian process, $z(\mathbf{s})$, which is the logarithm of the intensity measure. This transformation is required as the expected number of points cannot be negative. In order that this continuous process is adequately reflected a SPDE model can be used to approximate this spatially varying process. Specifically, the solution to the SPDE given in Equation (1.5), detailed by Lindgren et al. (2011), is a GRF. Thus, solving Equation (1.5) gives the required continuous process reflecting the intensity of the LGCP. One fundamental stage in obtaining the solution, is the triangulation—construction of the mesh—of the domain. This allows irregularly placed observations to be considered, using FEM, to interpolate the observation locations to the nearest (regular) grid point. This by the way of using deterministic basis functions and Gaussian weights—calculated from properties of the constructed mesh—to approximate the value of the Gaussian field at that location. Owing to this, it is required that the observation locations are mapped onto the triangulation (through the projector matrix), so that the neighbourhood structure of the data is specified, and accounted for in the modelling procedure. These two aforementioned stages are central to using a SPDE model for the latent field, hence, the sections above detail their specification. In addition, demonstrating the procedure in both `R-INLA` (Rue and Martino, 2007), and `lgcpSPDE` developed by the author of this thesis.

Throughout this chapter two example datasets—`lgcp2D`, and `terrorism`—are used, both included with the package `lgcpSPDE`. Respectively, they are simulated data, and data obtained from the GTD pertaining to recorded terrorism events (2002-2012). The former dataset is used to illustrate the sensitivity of the modelling technique to the triangulation. The latter data are used to illustrate the relevance of the GRF and its visualisation (Figure 4.4). The logarithm of the intensity measure shown in Figure 4.4 offers a foundation to understanding the mechanisms operating to proffer the spatial dispersion of terrorism events. Whilst, classically this structure—GRF—accounts for the remaining spatial structure in the data (in this case the only fixed effect is an intercept term) after the structure explained by

covariates has been accounted for. Thus, this structure evidently contains clue as to other mechanisms operating in proffering the observed point pattern.

In addition to the mechanism proffering the spatial dynamics of the point pattern there exists a process (conditional on the aforementioned spatial process) of which these characteristics are realisations. The point pattern is thus a marked point process. To illustrate this concept, consider a Show-jumping course[5] where the locations of the obstacles (fences) required to be jumped without fault by the competing horse and rider pair form a point pattern. Now, when either competitors or persons with adequate experience see the course—prior to anyone jumping it—they can identify obstacles which will potentially cause the competitors to obtain faults. These, problem obstacles will of course have been fashioned as such by the course builder. This associated degree of difficulty, is not purely down to height or width of the obstacle, but also location in relation to the other obstacles, colour of the obstacle, and approach to the obstacle. In the majority of cases the expertise of the aforementioned persons correctly identify the fences which will incur the most faults. Notwithstanding their beliefs, faults incurred at each obstacle throughout the competition gives information as to the difficulty level of the fence. Thus, any Show-jumping course is a realisation of a marked point pattern. Plainly, the intensity measure of this (finite) point process offers little useful information—the jumps are numbered. However, conditional on the obstacle locations the difficulty process operates in conjunction. Inferring this (mark) process—obviously specific to each individual course—is what is speculated at by the competitors etc. Through extending the methodology detailed in this chapter, particularly utilising multiple GRFs, will allow, through conditioning on (fence) point locations the latent process proffering the spatial distribution of the marks to be inferred. Moreover, it would also be possible to infer the dependency between the difficulty and location of obstacles. Thus, offering a general framework within which multiple mechanisms operating—perhaps in conjunction—to construct the intensity measure of a LGCP.

In this instance the additional information associated with the point locations (i.e., some continuous or discrete variable which accompanies the point location), are assumed to be realisations from a mark process denoted $M(\cdot)$. Specifically, letting $N(\cdot)$ be a point pattern, a realisation of some point process, then conditional on this process, there exists a model for $M(\cdot)$. Chapters 5, 6, and 7 extend the basic functionality discussed in the above sections. This, through developing methodology which enables further spatial processes operating in conjunction with $\Lambda(\mathbf{s})$, such as $\mathrm{M}(\mathbf{s})$, to be inferred.

[5]this section may have been written during the individual Show-jumping final of the 2016 Rio Olympics

Part IV

Shared stochastic structures in accounting for multiple dependencies

Chapter 5

The employ of shared stochastic structures in a class of extremely flexible latent Gaussian models

5.1 Introduction

The preceding chapters of this thesis discuss latent processes that are assumed to evoke the observed data structure of a point pattern. In particular, Chapters 2 and 3 focus on the concept of unobserved (homogeneously distributed) parent points which dictate the structure of the observed point pattern. Chapter 4 introduces the concept of a latent GRF that reflects the intensity measure of a point process—specifically a LGCP. Inferring the latent processes assumed to evoke the observed data structure is fundamental; either, purely to account for the spatial and or temporal auto-correlation in the data, or in order that a greater understanding of such latent mechanisms is gleaned.

Point pattern data are typically thought of as simply points in some d-dimensional space, however, frequently, they refer to physical objects with properties. These properties are referred to as marks (e.g., the size or the type of the object), and may provide an improved understanding of spatial structures if accounted for in the analysis (Illian et al., 2008). Marks may not only help explain the spatial structure formed by the objects, but may themselves exhibit a spatial structure of interest, which varies with the locations. Moreover, dependence between the marks and the point pattern may exist and be of interest. Therefore, modelling the point pattern in isolation would ignore the complexities inherent in the data structure, and neglect the dependence structures operating between the marks and the point locations. Thus, elements in a model representing such dependence are no longer nuisance parameters; their structure is relevant and interpretable in the context of the application. Hence, modelling approaches that consider dependence between marks and spatial point pattern further

increase the relevance of spatial point process models in applications.

More generally, inter variable dependency is of interest in many fields of study (Menezes et al., 2005). For example, in the field of epidemiology Knorr-Held and Best (2001), Kim et al. (2001), and Langford et al. (1999) jointly model several diseases. Likewise, ecologists may utilise similar methodology using shared stochastic structures to capture the relationships among multiple species. Specifically because within an ecosystem few components are independent due to the multitude of complex interactions. Thus, in order to infer, or simply account for such dependencies, a more unifying modelling approach over and above the assumption of independence is required.

This chapter chiefly focuses on utilising latent structures—in particular, GRFs—in order to facilitate the formulation and fitting of a marked LGCP model. Multiple latent structures are utilised in the proposed modelling framework to capture the inherent structures and dependencies in marked point patterns. This chapter specifically focuses on their relevance in application to data from the field of geology.

Before moving onto considering this, Section 5.1.1 introduces the concept of shared processes, and dependence between variables. This through extending the ideas discussed in Chapter 4. Section 5.1.2 details some functionality of the `R` package `lgcpSPDE` in simulating and fitting a marked LGCP model. In addition, that section demonstrates estimator performance through detailing a small simulation study, as well as illustrating the considered modelling ideas and concepts. The prime focus of this chapter is presented in Section 5.2 where the flexibility and versatility of the model class developed herein is detailed through an application to geological data.

It should be noted that models discussed in this chapter were fitted using the INLA–SPDE methodology (Chapter 4). The steps required to fit each model are not detailed; purely the model construction is discussed and inference drawn. In addition further functionality of the package `lgcpSPDE` is introduced.

5.1.1 Shared processes

In order to introduce the concept of a joint model one should consider two random variables—$N_\cdot$ and $Y_\cdot$—which are assumed to be a realisation of some spatial or spatio-temporal stochastic process. Here the subscript $\cdot$ is either a spatial or spatio-temporal index. In particular, let their respective means, μ_N & μ_Y, be given by

$$\begin{aligned} \eta(\mu_Y) &= g(\cdot), && \text{and,} \\ \nu(\mu_N) &= \beta_0 + \beta\, g(\cdot) + h(\cdot), \end{aligned} \tag{5.1}$$

for some given link functions $\eta(\cdot)$ & $\nu(\cdot)$. Here β_0 is an intercept term. Most notably however, $g(\cdot)$ and $h(\cdot)$ are spatial or spatio-temporal processes which capture the inter variable dependence. Moreover, the parameter β—termed the interaction parameter—describes the

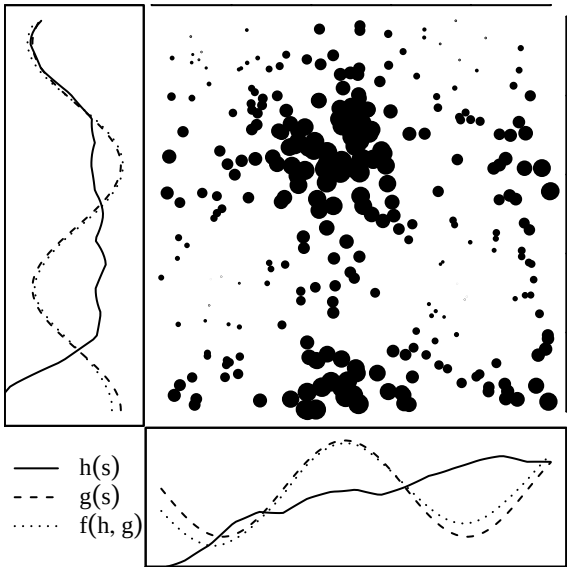


Figure 5.1: Plot showing a marked point pattern simulated in the unit square. The size of the plotting characters reflect the magnitude of each point's mark. Along each axis are plots showing the average value of the processes contributing to the point pattern. Solid line $h(\mathbf{s})$; process governing the intensity of the point locations conditional on the mark process. Dashed line ($g(\mathbf{s})$); process governing the mark. Dotted line f(h, g); a linear combination of the afore mentioned processes. It is perhaps clear from the pattern that the intensity of seemingly oscillates in both the x and y directions. Thus, it is clear that there exists dependence between the spatial distribution of the point pattern and mark. However, the plots adjacent the axes show that there also exists some other structure conditional on the mark process.

contribution of the process $g(\cdot)$ to the structure inherent in $N.$. Letting $\phi = (\kappa, \sigma^2)$ be the parameters of the processes $g(\cdot)$ & $h(\cdot)$, the parameters of the likelihood, $L(\boldsymbol{\theta})$, for each random variable $Y.$ and $N.$ are then given by $L(\phi)$ and $L(\beta_0, \beta, \phi)$ respectively. Now, if $N.$ and $Y.$ were independent then $\beta = 0$, and the structure of $N.$ would be wholly attributed to $h(\cdot)$. However, if there were dependence between $N.$ and $Y.$—β would thence be non-zero—$h(\cdot)$ would capture the latent spatial or spatio-temporal structure unique to $N.$ conditional on the structure of $Y.$.

Figure 5.1 shows a realisation of a marked point process[1]. Here $Y.$ is a realisation of a marked LGCP, with a mark given by $N.$. This simulated pattern clearly illustrates strong dependence between the point locations and the size of the mark, (i.e., the intensity of the process oscillates in both the x and y direction, as does the mark's magnitude, which is illustrated by the size of the plotting characters). Thus, it is clear that there exists dependence between the spatial distribution of the point pattern and mark. Using the construction above, this mark structure (illustrated be the dashed lines in the plots adjacent the axis in Figure 5.1) is represented by $g(\cdot)$. This structure is shared between both processes, and the parameter β—the interaction parameter—captures the dependence between the point locations and the mark.

Based on the concept illustrated by Figure 5.1 and Equation (5.1) above, the following

[1]This simulated point pattern is obtained using the function `rlgcpspde()` function in package `lgcpSPDE`

section details constructing a general class of multivariate latent Gaussian models which incorporate inter mark dependence. It should be noted that such a modelling framework can be utilised to model seemingly very different data structures. One method of estimating the interaction between the spatial distribution of points and the point pattern marks would be to consider the marks as covariates and treat them as fixed effects. However, this may oversimplify the complexity of the inherent dependency structures, as any auto-correlation between marks is ignored, and under assumptions of the approach taken here, impractical, as covariates must be observable everywhere. Moreover, this may result in different models and therefore interpretations in terms of causality. Here, latent structures are used to infer the different types of dependence and interaction. While such an approach has been explored before by Illian et al. (2012b), it is herein extended to account for interaction between marks. In addition one can easily generalise such an approach to include a temporal dimension.

In their simplest role the random fields mop up the spatial structure caused by unobserved covariates (Møller and Waagepetersen, 2007). However, a structure which simply reflects the spatially varying intensity of objects or events is not necessarily of interest. Hence, this chapter focuses on the development of point process methodology, which(i) extends and generalises existing methodology, adequately capturing the complexity inherent in the marked point pattern data, (ii) presents modelling results that are both meaningful and useful to the user community, and finally, (iii) are both extremely flexible and relatively straightforward to fit encouraging their use amongst non-statisticians.

Section 5.2 discusses marked point process models which utilise shared stochastic structures. Section 5.1.2 gives a brief simulated example—where data is simulated from the model given by Equation (5.1)—in order that estimator performance may be assessed. Thereon, Section 5.3 details the methods discussed above and its application in the field of geology.

5.1.2 A Simulated example

This section demonstrates some extended functionality of the package `lgcpSPDE` (introduced in Chapter 4). Specifically, in simulating and fitting marked LGCP models of the form detailed in Equation (5.1).

Simulating a marked LGCP–the `rlgcpspde()` function.

The function `rlgcpspde()` can be used to simulate data of the form given in Equation (5.1). In addition to the arguments discussed in Chapter 4, the function `rlgcpspde()` has arguments which enable the simulation of a marked point process. The three main additional arguments to consider are,

- `mu`: a numeric scalar specifying the mean intensity of the LGCP (β_0 in Equation (5.1))
- `mark`: Logical, if TRUE a marked point pattern is simulated.

- `beta`: A numeric scalar, the interaction parameter describing the dependence between the mark and point locations. By default this is 1, if `mark = TRUE`
- `mark.function`: A function of 2D spatial coordinates (`(x,y)`) or a 2D pixel image which describes the spatial process specific to the mark, by default this is `function(x,y){cos(x) - sin(y)}`

The object returned after calling the function—as `marked_pp` in R Example 5.1.1—is a $3 \times m$ matrix; the first two columns the m coordinates of the simulated LGCP the third the associated mark.

The code in R Example 5.1.1 illustrates how to simulate a marked LGCP; the spatial structure of the mark process is supplied through the argument `mark.function = mark.im` where `mark.im` is a pixel image of `class im`. In addition, the object `spatial.polygon` constructed in R Example 4.3.3—a 10×10 square specifying the domain—is required.

R Example 5.1.1.

```
marked_pp <- rlgcpspde(spatial.polygon = spatial.polygon,
                       mesh.pars = c(min = 1, cutoff = 1.2), kappa = 0.5,
                       mu = 2,
                       mark = TRUE, beta = 1, mark.function = mark.im ,
                       seed = 1)
```

Fitting a marked LGCP–the `fit.marked.lgcp()` function.

The fitting of a marked LGCP model—using INLA and a SPDE model for the latent fields—is achieved via the `fit.marked.lgcp()` function. The object returned is an `inla` object providing information as to the fitted marked point process model which can be extracted through utility functions (as per `R-INLA`). The structure of the fitted model is as specified in Section 5.1—and can either be spatial or spatio-temporal—where the random fields are named; `field.pp`, `field.mark`, and `copy.field`, referring to $h(\cdot)$, $g(\cdot)$ and $\beta\, h(\cdot)$ respectively. The main arguments of `mark.pp.fit()` are,

- `mesh`: A Delauney triangulation of the area, specifically, either an object returned by `inla.mesh.2d()`, or the wrapper function discussed in Section 4 `make.mesh()`.
- `locs`: A—$m \times 2$—matrix containing the (m) observed location of the point pattern.
- `mark`: A vector of length m of mark values referring to each point location.
- `mark.family`: A character string specifying the assumed likelihood for mark, by default this is `"gaussian"`.

Table 5.1: Estimated biases, standard deviations, and root mean square errors for estimators of interest: interaction parameter β, intercept term for the LGCP β_0, and parameters (κ_N, σ_N^2) and (κ_Y, σ_Y^2) of each GRF $h(s)$ and $g(s)$ respectively All are given as percentages of the true underlying parameter values.

Estimate	Bias(%)	SD(%)	RMSE(%)
$\hat{\beta}$	-9.4895	5.5739	11.0026
$\hat{\beta}_0$	0.5178	8.2488	8.2568
$\hat{\kappa}_N$	4.0280	40.7978	40.9555
$\hat{\kappa}_Y$	1.2366	3.2845	3.5065
$\hat{\sigma^2}_N$	49.6953	111.7759	122.2231
$\hat{\sigma^2}_Y$	5.4108	6.1241	8.1674

Moreover, the user may set the prior on the interaction parameter describing the dependence between the mark and point locations, this by default is $N(0, 10)$, and can be changed through changing the default argument `hyper = list(theta = list(prior = 'normal', param = c(0,10)))`. The user can fit a spatio-temporal marked point process model through specifying the argument `t.index` to be a numeric vector of length m specifying each time index of each observation (i.e., $1 \cdots T$). If a spatio-temporal model is fitted the user may also specify the prior on the AR(1) temporal coefficient.

To illustrate fitting a marked point process model the simulated data contained in the object `marked_pp` is used. Executing the code in R Example 5.1.2 (which uses functionality of the package `lgcpSPDE`) will fit a marked point process model.

R Example 5.1.2.

```
fit <- fit.marked.lgcp(mesh = mesh, locs = marked_pp[,1:2],
                    mark = marked_pp[,3], mark.family = "gaussian",
                    verbose=TRUE,
                    hyper = list(theta = list(prior = "normal",
                                              param = c(0, 10))))
```

5.1.3 Estimator performance

In order to asses estimator performance 500 datasets were simulated from the model given by Equation (5.1) with the same parameter values as those used to simulate the data `marked_pp` in R Example 5.1.1. Each was analysed using the function `fit.marked.lgcp()`.

The parameters of the model are: the interaction parameter β, intercept term for the LGCP β_0, and parameters (κ_N, σ_N^2) and (κ_Y, σ_Y^2) of each GRF, $h(s)$ and $g(s)$ respectively. Figure 5.2 shows the difference of each posterior parameter value from the true value, as a percentage. Moreover, Table 5.1 shows estimated biases, standard deviations, and root mean

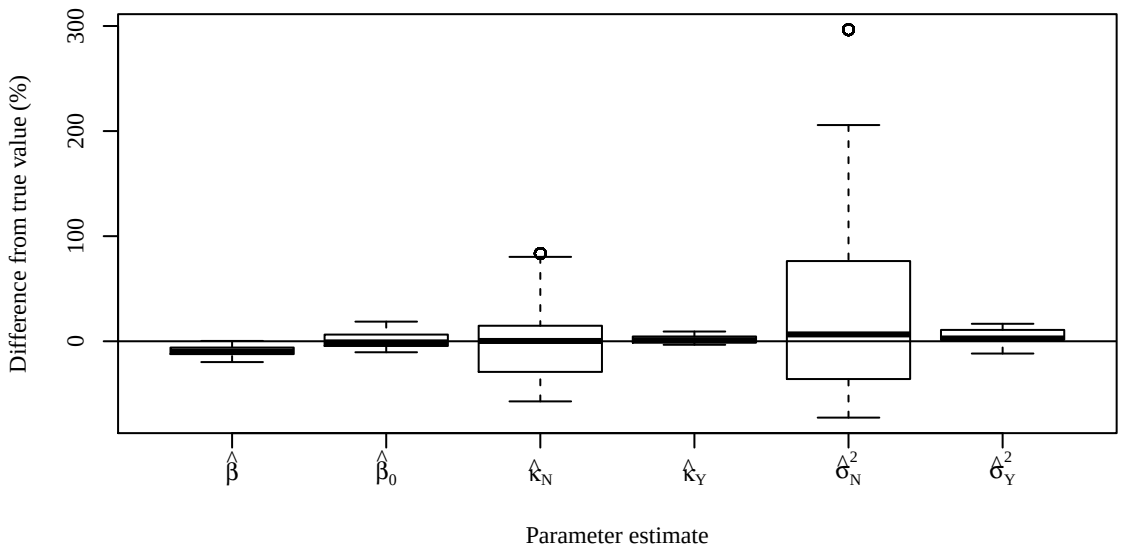


Figure 5.2: Boxplots showing the percentage differences between the true and estimated parameter values from the model in Equation (5.1). Those being the interaction parameter β, intercept term for the LGCP β_0, and parameters (κ_N, σ^2_N) and (κ_Y, σ^2_Y) of each GRF $h(s)$ and $g(s)$ respectively.

square errors for estimators of interest, as percentages of the true underlying parameter value. The interaction parameter estimator is slightly negatively biased. The variance estimator for the random field $h(s)$ is positively biased, and has large SD(%). This, perhaps due to the inclusion of an additional random field in the model.

5.2 In practice: the roles of a latent field

In the following section, a joint modelling methodology is applied in modelling the spatial dynamics inherent in marked point pattern data (Myllymäki and Penttinen, 2009; Ho and Stoyan, 2008). In particular, the modelling framework proposed is based on the assumption of a marked LGCP (Møller et al., 1998; Moller and Waagepetersen, 2004; Brix and Diggle, 2001). The GRF's function is to account for the surplus spatial structure which remains after the variation due to the considered covariates has been accounted for. Specifically, the remaining unexplained spatial structure in the data which may proffer information as to unconsidered or unknown mechanisms aiding generation of the observed morphology. Through modelling a point pattern and the associated marks concurrently the uncertainty relating to each model component is propagated and attributed correctly to each component. The interpretation of the inferred latent stochastic structure can be thought of as a process driving the observed marks conditional on the point locations.

In general, the random fields in such models take on very different roles dependent on the application. It should be noted that due to the flexibility of such a joint modelling approach the random fields can be used to reflect different types of mechanisms. Thus, providing very

different types of answers and insights into the underlying mechanisms that generate the modelled pattern. To illustrate this the presented application comes from the field of geology (earthquake prediction). Particular emphasis is put on the class of model proposed. This as through utilising multiple GRFs in such a joint modelling approach culminates in moving away from regarding the random field merely as a way of accounting for unexplained spatial structures. Specifically, considering more complex models in which random fields represent interpretable structures increases the relevance of the models developed herein.

The following section (Section 5.3) considers data from the field of geology (earthquake prediction). The dataset details the locations and magnitudes of recorded earthquakes worldwide in 2013[2]. The locations in the dataset are considered to be a realisations of a marked LGCP. The modelling framework proposed in this context utilises shared Gaussian random fields to model the dependencies between marks and point locations.

5.3 Application to geological data: a marked point process for earthquake location and magnitude

The data considered in this section refer to the estimated earthquake epicentres, and magnitudes of earthquakes recorded in 2013. The earthquake epicentres are assumed to be a realisation of a marked LGCP with the associated marks being their recorded magnitude measured on the Richter scale (see Figure 5.3). The recorded Richter magnitudes range between 4.98 (light) to 9.60 (great), and were standardised for this analysis (purely for computational efficiency). The Richter magnitude of an earthquake is determined from the logarithm of the amplitude of waves recorded by seismographs.

Without detailing the vast geo-physical reasons for the occurrence of earthquakes, the lay-person understands that the seismic waves (which make the ground vibrate), are caused by the release of energy resulting from underground rock breaking along a fault line (cracks in the tectonic plates below Earth's crust). It is well established the likely regions where one would expect earthquakes to occur, broadly speaking these correspond to the fault lines of the tectonic plates. This structure is evident in the patterning shown by the recorded earthquakes in Figure 5.3. There are of course far more complex processes associated with earthquake locations, they are not exclusively the result of aftershock from earthquakes, and geodetic strain rate. Conditional on the known epicentre locations (points) there may be some additional process which is specific to the magnitude (mark) of the earthquakes.

Let the random variable $N(\Omega)$ be the point pattern pertaining to the earthquake epicentres, then assuming that $N(\Omega)$ is a realisation of a LGCP (as described in Section 1.2.2). Therefore, $N(\Omega) \sim \text{Poisson}\left(\int_\Omega \lambda(\mathbf{s})\mathrm{d}s\right)$, where $\Lambda(\mathbf{s}) = \exp(z(\mathbf{s})) = \int_\Omega \lambda(\mathbf{s})\mathrm{d}s$ is the intensity measure of the process. Thus, in the context discussed here, $\exp(z(\mathbf{s}))$ represents the spatial

[2] The data are available from the International Seismological centre (`http://www.isc.ac.uk/`).

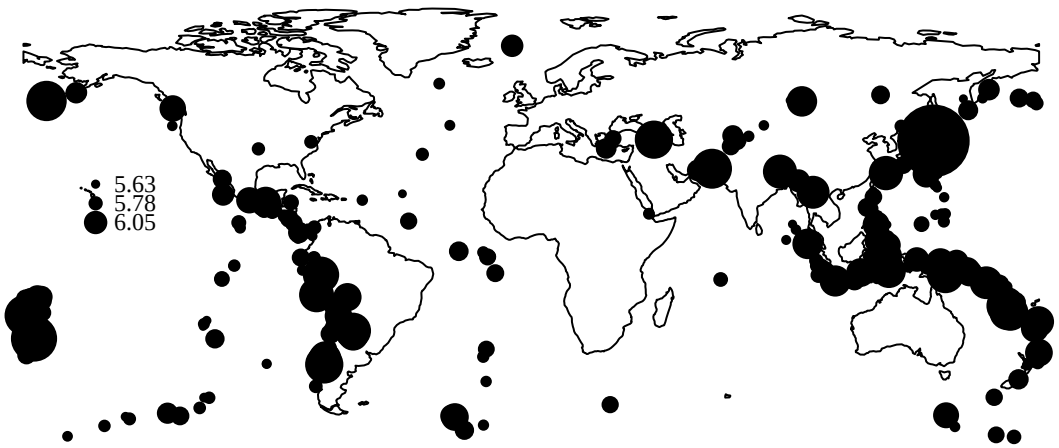


Figure 5.3: Epicentre locations worldwide of each recorded earthquake in 2013 assumed to be a realisation of a LGCP. The size of each plotting character reflects the magnitude of each earthquake. The evident structure visible in the pattern is linked to the Earth's fault lines.

distribution of earthquake locations, where $z(\mathbf{s})$ is a GRF. Now, the intensity measure $\Lambda(\mathbf{s})$ is not necessarily of interest—for reasons suggested above—yet conditional on this process, there exists a model for the magnitude (mark), $M(\mathbf{s})$. Due to this dependence between the point pattern and mark, a shared GRF is used to reflect the dependence between the spatial pattern and the magnitude. This dependency itself might be of interest. The full joint model—similar to the model in Equation (5.1)—is given by

$$\begin{aligned} E(N(\Omega)) &= \exp(\alpha + z(\mathbf{s})), \\ E(M(\mathbf{s})) &= \beta_0 + z(\mathbf{s}) + z_1(\mathbf{s}), \end{aligned} \tag{5.2}$$

where $M(\mathbf{s}) \sim N(\beta_0 + \mathbf{z}(\mathbf{s}) + \mathbf{z_1}(\mathbf{s}), \sigma_M^2)$ (noting the magnitudes are standardised for this analysis). Both α and β_0 are intercept terms, and σ_M^2 is the mark variance. Both $\mathbf{z}(\mathbf{s})$ and $\mathbf{z_1}(\mathbf{s})$ are zero mean Gaussian random fields. The parameter vector is thus $\boldsymbol{\theta} = (\alpha, \beta_0, \phi_z, \phi_{z1}, \sigma_M^2)$ where both $\phi_z, \phi_{z1} = (\kappa, \sigma^2)$ are parameters of each GRF. Here, $\mathbf{z}(\mathbf{s})$ is shared between the mark and the point pattern, capturing the point-mark dependence. That is, the joint model defined above, using two likelihoods, Poisson for the point pattern and Gaussian for the mark, attributes all the spatial dynamics of point locations to random effects, then assumes that this structure is shared with the mark. Hence, the mark specific additional stochastic structure—$\mathbf{z_1}(\mathbf{s})$—captures the mark specific spatial structure, conditional on the point

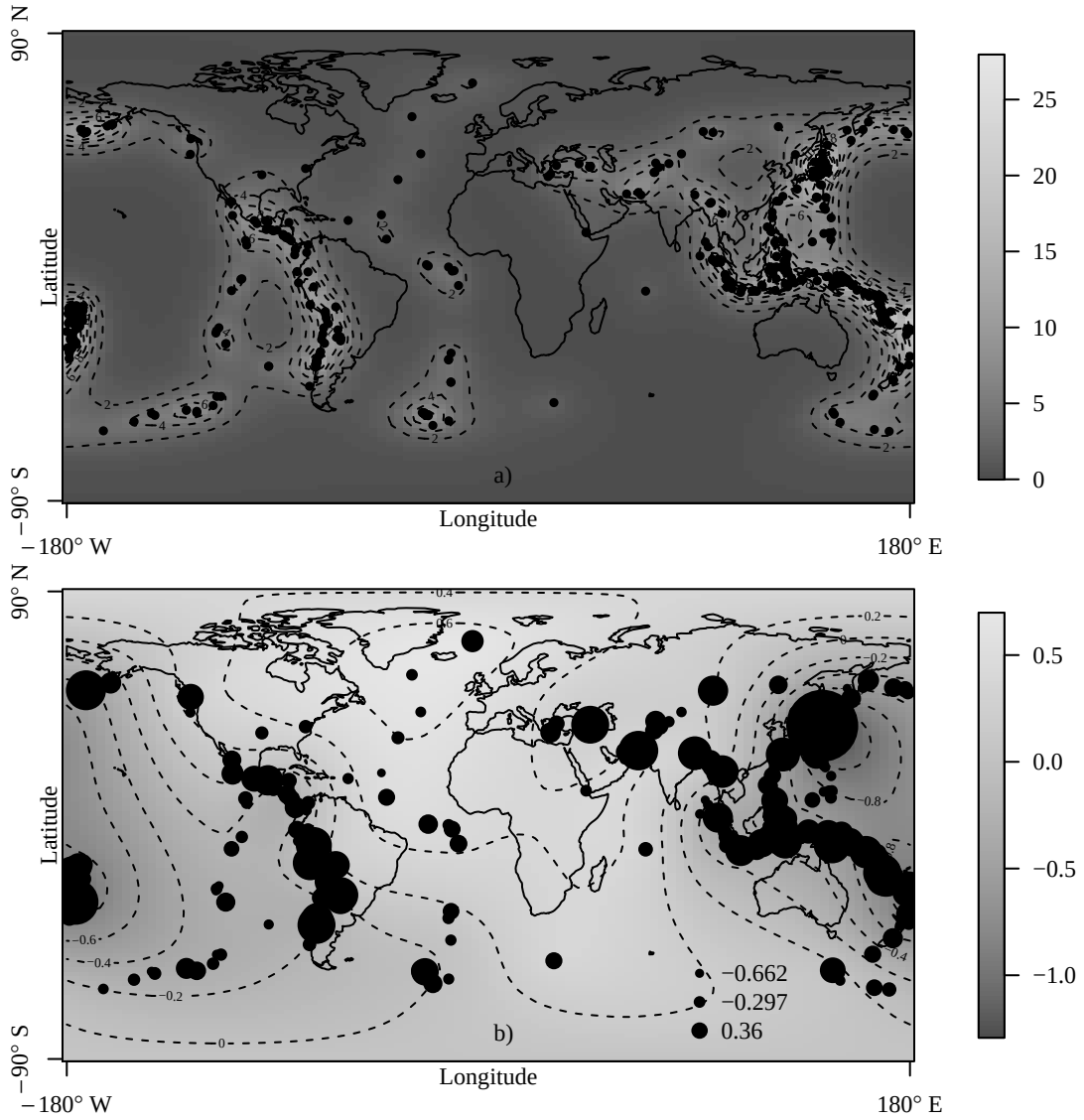


Figure 5.4: Plots of the posterior means of each GRF in Equation (5.2). Plot a) shows on the response scale (i.e., intensity) the field $z(\mathbf{s})$; plot b) shows the field $z_1(\mathbf{s})$ (i.e., the scaled earthquake magnitude conditional on earthquake locations). Each plot shows the epicentre locations of the earthquakes; the size of the plotting characters in the latter plot represent the scaled magnitudes.

locations.

5.3.1 Results and Inference

As noted above, inferring the spatial auto-correlation inherent to point location is not necessarily of interest, as this stochastic structure is essentially the point intensity (spatially varying intensity measure). This is somewhat obvious before any model fitting is carried out. As is clear from Plot a) in Figure 5.4 it adequately reflects the spatial distribution of the observed points. That is, this structure reflects the intensity measure of the point process begetting earthquake epicentres. In formally assessing the validity of the assumption that

Table 5.2: Posterior means, standard errors, and 95% credible intervals for the parameters α, β_0, and σ^2_M (Equation (5.2)).

			Quantile of Order	
Parameter	Mean	SD	2.5%	97.5%
α	-1.9555	0.9269	-3.8475	-0.0269
β_0	-3.5068	0.5289	-4.6171	-2.4802
σ^2_M	0.9875	0.0606	0.8801	1.1137

these points are a realisation of a LGCP is touched shown by Figure 5.5. Here a functional summary characteristic (J-function) of the observed point pattern is compared to those of 100 simulated LGCPs of a similar structure. That is, point patterns were simulated based on the estimated intensity surface of the LGCP in Equation (5.2) and their structure compared to that of the observed point pattern of earthquakes. The J-function is a combination of the nearest neighbour distribution function and the empty space function (see Section 1.2.1). Recall from Section 1.2.1 that for a homogeneous Poisson process $J(r) = 1$ as $G(r) = H(r)$. Note also, that for a cluster process, $J(r)$ is ≤ 1 for $H(r) < 1$. This is because the distances between points of the process tend to be shorter than distances between the origin and the nearest cluster (i.e., $1 - G(r) \leq 1 - H(r)$). From Figure 5.5 it seems that the earthquake locations are clearly clustered ($\hat{J}(r) \leq 1$), however they are unlikely to be a realisation of a LGCP. This perhaps was to be expected, as due to the link between fault lines and earthquakes, the clustering evident in earthquake location is not random across space but along a network. Figure 5.5 indicates that, at shorter distances, the J-function for the earthquake data falls at a steeper rate than those for LGCPs simulated from the fitted model. This indicates that the quotient of $1 - G(r)$ and $1 - H(r)$ for the earthquake data decays towards 0 at a faster rate than that of a LGCP with the same inhomogeneous mean. This is to be expected: from the pattern shown in Figure 5.4, the clusters of points are not randomly distributed, hence the distances between points (both from some test location and a typical point) are shorter than for a typical LGCP.

Table 5.2 gives the posterior means, standard errors, and 95% credible intervals for the parameters α, β_0, and σ^2_M with respect to the joint LGCP discussed above. Both intercept parameters, α and β_0, referring to the point process and mark respectively are negative (on the link scale). These can be thought of as the average intensity or mark value respectively conditional on the other attributes of the linear predictors. Perhaps of most interest however is the residual spatial structure, once the dependence between point locations and the mark is accounted for. Residual structure has been used by Schoenberg (2003), and Ogata (1988) to asses model fit. In the latter case the possibility of using such analysis in order to identify features of the data not accounted for in the model (focusing on seismic quiescence) is discussed. The approach discussed here focuses on inferring the features and processes not

accounted for in the model through the residual structure. It might be tempting—when using a point process approach—to think that no spatial structure remained once the intensity measure of the process had been accounted for. That is, the spatial structure of magnitude being wholly accounted for by the drivers of the earthquakes themselves. However, Plot b) in Figure 5.4 shows some spatial structure which may be of interest when identifying unconsidered processes operating over space. This structure suggests that earthquakes in Europe and north-east America seem to be stronger than what would be expected simply given the earthquake locations. Specifically, conditional on the intensity measure of the earthquake's epicentre locations there is seemingly an additional process operating of which the observed magnitudes are a realisation.

As mentioned above one might not consider a LGCP to be an appropriate model for the locations of earthquakes. Despite this, the mark (magnitude) process inferred as a result of the joint model illustrates the generality of the proposed model. To illustrate the type of inference that might be gleaned from the proposed modelling framework the interpretation of the estimated mark process is put into the context of the data discussed in this chapter. This chapter's aim is not driven by the requirement of fitting a wholly appropriate model to this data. If this were the case one perhaps might want to consider a different spatial correlation function, although if this were the case the SPDE case would not be possible. To appropriately model earthquake locations on a large scale across multiple tectonic plates, it is necessary to somehow incorporate spatial dependence due to the fault line network.

As far as the author of this thesis is aware little work has been done assessing the residual structure of earthquake models worldwide. Both Schoenberg (2003), and Ogata (1988) focus on a much smaller spatial scale (i.e., California and Japan), concentrating on changes over time of the residual structure. Therefore interpretation of the estimated mark process—Figure 5.4, Plot b)—is pure conjecture. However, this spatially varying process may be linked to the crustal strain that the Earth is subjected to during the occurrence of an earthquake. Composition, mineralogy, pressure, and temperature of the Earth's crust all affect its ability to carry and transmit stress, thence, dictate how the crust will deform. The waves' amplitudes are affected by the physical properties of the Earth's crust which clearly vary spatially. Thence, one might suppose that the mark process inferred in Figure 5.4 (Plot b)) is reflecting the physical properties of the Earth's crust, and thus, the seismic wave propagation.

5.4 Discussion

This chapter has detailed development of methodology which facilitates the analysis of complex dependencies inherent in marked point pattern datasets. Many processes occurring at discrete locations in either space or in both space and time can be formulated in a way amenable to the approach presented here. This chapter has chiefly focused on highlighting

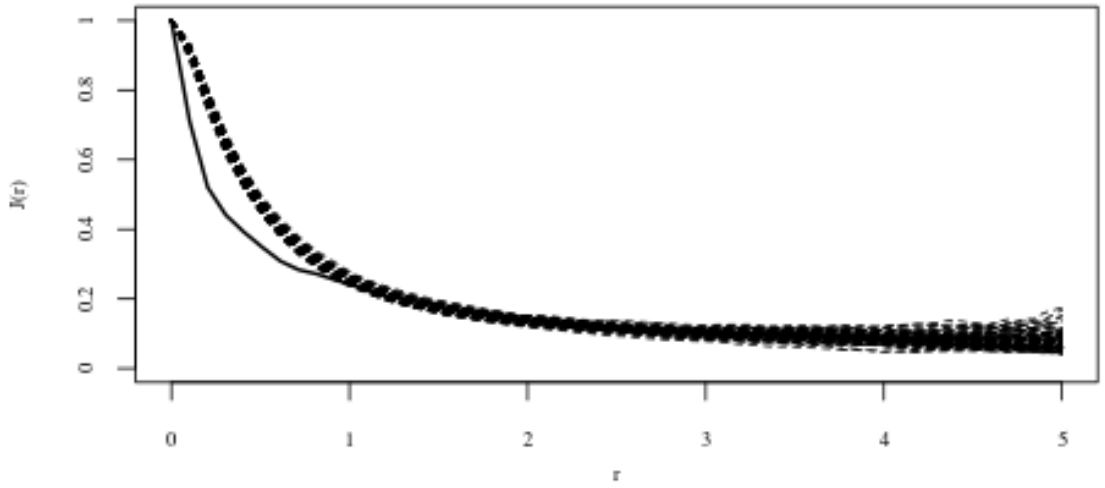


Figure 5.5: Plot comparing the J function summary characteristic for the point pattern formed by the earthquake epicentres (solid line) and 100 simulated LGCPs (dashed lines) at distances r, in degrees. Each simulated point pattern was simulated with the estimated intensity of the earthquake point pattern from model Equation (5.2).

how, through the combination of INLA and the SPDE approach, spatial structures and mark dependence can be accounted for. Thus, enabling one to infer the residual spatial structure inherent in the data. In addition, inferring the residual processes conditional on the intensity of the observed point pattern offers insight into the mechanisms specific to the spatial distribution of the characteristics of the point pattern.

Moreover, the inclusion of the spatial random effects representing the latent structures in the data in a joint modelling framework enables one to infer the remaining unexplained spatial structure of the marks conditional on the point locations. Thus, offering information as to the unconsidered processes unique to the marks, as well as the correlation amongst marks in a multiple marked point process. In practice, the mechanisms driving the observed marks of a point pattern have very different interpretations depending on their nature, despite the mathematical construction of the models being similar. As such, even though the modelling procedure theoretically is the same, in reality, the interpretation of the modelling output is very much context dependant, as the role of the stochastic structures differs in each context. It should be noted that the roles of the spatially structured random fields within this framework are an extension of their traditional roles. Specifically, in addition to reflecting the spatially varying intensity of objects or events, the spatial processes that are mark specific conditional on their locations can be inferred. It is also paramount to note that the practical interpretation of these stochastic structures is specific to the particular application, and requires comprehensive dialogue between the statistician and practitioner from the field of study. This problem is discussed by (Illian and Burslem, 2015).

The proposed approach has been illustrated through modelling the dependence between

point locations, and associated marks in the field of geology. The spatial structure specific earthquake magnitude conditional on earthquake location is estimated enabling areas of high magnitude earthquakes conditional on their intensity to be inferred.

In summary, based on the assumption of a latent Gaussian class of model being appropriate for the type of data considered herein, this chapter demonstrated that the proposed methodology can account for inherent dependencies in many datasets. In addition, the methodology simultaneously assessed the residual processes conditional on the locations of events proffering the spatial distribution of marks. Through implementing a joint model for the point process and its mark the uncertainty relating to each component is propagated through the modelling approach.

This chapter concentrates on applying the methodology to marked point pattern data that focuses on the spatial pattern formed by events in space. However this can easily be adjusted for geostatistical data (i.e., where a spatially continuous model is fitted to measurements taken at a finite number of locations). The flexibility of the developed general class of joint marked point pattern models is illustrated in the following chapter (Chapter 6) where the inherent complex relationships in ecological data are inferred.

Section 6.2, utilising similar mathematical construction of the joint latent Gaussian models discussed herein, applies the methodology to multi-species data with a semi continuous response. Although the data is spatial—and temporally referenced—it is not considered as a realisation of a point process. That is, the multi-species data (Section 6.2) is modelled as geostatistical data observed at a set of locations throughout the UK. The reader may note that the data discussed in Section 6.3 is of a similar structure, that is, spatially referenced observations pertaining to animal presence. Yet, Section 6.3 focuses on inferring species' spatial distribution conditional on their habitat dependence.

Chapter 6

The use of multiple latent Gaussian fields in inferring ecological interactions

6.1 Introduction

As in many other physical world phenomena ecological systems are driven by both dynamic processes, and noise processes (Wood, 2010). Typically, models that contain some stochasticity are used to account for the noise process. Yet, these—noise—structures can also be utilised to inform regarding the inherent dynamic processes in the system. Consider the field of ecology, rarely do processes occur in isolation; regard for example the mechanisms which facilitate the co-occurrence of different species (in the same spatial area at the same time). Treating the species in isolation would in many instances be unrealistic, as, this would ignore the inherent dependencies. Moreover, the inherent dependence structures may be the very mechanisms of interest. It is these latent structures—in an ecological context—which are considered in this chapter.

All previous chapters discuss the incorporation of latent components in the fitting of point process models, Chapters 2 and 3 through unobserved parent points resulting form a homogeneous Poisson process, Chapters 4 and 5 through a latent GRF. In addition to the incorporation of a latent structure in point process models, GRFs are commonly used in the modelling of other types of spatial data (e.g., spatially continuous geostatistical data, see Section 1.1). In a slight change to the main focus of this thesis—the analysis of point pattern data through the use of NSPPs, Chapter 2, void processes, Chapter 3, or LGCPs Chapters 4 and 5—this chapter considers utilising GRFs in the modelling of geostatistical data (pertaining to a semi-continuous and binary response).

Shared GRFs are utilised in this chapter to represent multivariate underlying processes

inherent in the data. This through essentially extending the linked model proposed by Diggle and Milne (1983), which sets two fields proportional to one another yielding a class of multivariate GRFs similar to those termed linear models of coregionalisation (LMC). Extensions of this approach are based on incorporating additional and shared random fields, the inclusion of which enables the interaction among multiple responses to be inferred. The class of multivariate latent Gaussian models developed in this chapter are fitted using the INLA—SPDE approach detailed in Sections 1.3.1 and 1.3.2 (and Chapter 4).

Section 6.2 details an application to multi-species avian data; the interactions, in space and over time, between species in the same and in different trophic levels are inferred. Section 6.3 details an application to habitat data; conditional on habitat the probability of presence of *Athene noctua* (little owl) is inferred. Both applications demonstrate how, through the use of stochastic structures in the modelling procedure, interactions intrinsic in ecological data may be inferred.

Multi-type latent Gaussian models in an ecological context

Chapter 5 outlined the use of multiple latent stochastic structures—GRFs—in fitting marked point process models. The models detailed in Chapter 5 were of the same mathematical construction, and thus, resulted in a general modelling framework for multiple dependent variables. This chapter extends such methodology to infer interactions in an ecological context. Such interactions are, in the first given example, modelled over time. Thus, offering scope in assessing the inherent complex dependencies over time of multiple dependent spatial variables.

The main interest in the motivating multi-species ecological data considered in Section 6.2 is not only in each species' spatial distribution, but also assessing how their spatial distribution changes over time. This as, the interaction between species within and between trophic levels often varies in space and over time; understanding these inter- and intra-specific interactions is imperative to discern the complex ecological relationships present.

It may be presumed that species in close proximity to one another potentially compete for similar resources. Thus, modelling individual species ignoring any effects of this competition, and hence relevant biotic interactions, while only focusing on abiotic interactions, would oversimplify the underlying dynamics (Boulangeat et al., 2012). One method of taking interactions among species into account is to model the dynamics of one species as the response variable of interest, and include the other species' density as a fixed effect. However, the uncertainty relating to each model component is neither propagated nor attributed correctly to each component. Including components in a model which should accommodate errors due to any source of variability as fixed components makes parameter estimates and thence predictions unrealistic (Clark, 2003). Moreover, such an approach assumes specific directionality implying that one is able to choose a focal species, which may be sensible in

some contexts where background knowledge may be used to justify this simplification, such as in Illian et al. (2009); Högmander and Särkkä (1999). However, in complex ecosystems the dependence relationships are rarely this simple (Dray et al., 2012). Thus, to provide a realistic modelling approach and to enable the relevant dependence structures to be inferred a multivariate approach is herein proposed.

This chapter, in a similar vein to the methodology discussed in Chapter 5, considers applications to geostatistical data from ecological contexts. The following sections respectively consider multi-species data referring to the British Trust for Ornithology's Garden Bird Feeding Survey (GBFS), and data relating to *A. noctua*s, and their possible nesting sites. Work discussed in Section 6.2 is detailed in the submitted manuscript *Jones-Todd, C., Swallow, B., Illian, J., & Toms, M* "A spatio-temporal multi-species model of a semi-continuous response" *Journal of the Royal Statistical Society. Series C (Applied).* Section 6.2 describes work undertaken by the author in the development of the model discussed, however, credit should be given to Ben Swallow for the ecological interpretation, an in-depth version of which is not presented here.

6.2 A spatio-temporal multi-species model of a semi-continuous response

This section considers multi-species data, specifically, the British Trust for Ornithology's GBFS [1] data. The raw data collected by the British Trust for Ornithology (BTO) are count data, collected at approximately 200 sites per year across the UK over a 36 year period, 1970-2005 (black plotting characters in Figure 6.1). The maximum count of each species seen feeding at food provided at the surveyed garden are noted in each of up to 26 weeks each winter spanning the months October to March. Herein annual averages are considered. That is, averages across weeks giving a mean of weekly maxima for each site-year combination. Thus, the data analysed in this section relate to what is assumed to be a continuous variable with a non-zero probability of obtaining exact zeros, sometimes referred to as a semi-continuous response variable (Aitchison, 1955). It should be noted that technically the distribution of the means is still a discrete distribution, however following methodology of Swallow et al. (2016) it is assumed that a continuous approximation is appropriate. Previous modelling approaches for such semi-continuous data introduce the use of a delta-gamma model, used in this chapter, that jointly models the binary and continuous sections of the response (Foster and Bravington, 2013).

The majority of multi-species approaches usually explore either singularly (i) a predator-prey relationship (Garneau et al., 2007), or (ii) co-occurrence of sympatric species (see Schweiger et al. (2012)) . However, in nature these different types of relationships rarely

[1] `http://www.bto.org/volunteer-surveys/gbfs`

Figure 6.1: Triangulation of the domain *(UK)* with 1596 verticies, and the extended mesh beyond the domain. Black circular dots are the nodes of the triangulation, black dots are the 693 sites. The black line represents the Spatial Polygon describing the border of the UK.

occur in isolation, and instead operate simultaneously. In order that the complexities of dependence relationships are adequately captured a multivariate approach is required, where shared spatio-temporal random effects give invaluable insight into the interaction between each species. That is, biotic interactions betwixt species potentially competing for similar resources can be considered. This is fundamental when considering complex ecosystems, as focusing purely on abiotic interactions, ignoring biotic dependencies, may drastically over-simplify the underlying dynamics of the system (Boulangeat et al., 2012). The benefit of a multivariate approach is that it allows uncertainty to be correctly attributed, and propagated. This as, typically biotic relationships assume specific directionality implying that there is a focal species (Illian et al., 2009; Högmander and Särkkä, 1999). Hence, the non-focal species are included as fixed effects, where in fact they should accommodate errors, thus, making parameter estimates and thence predictions unrealistic (Clark, 2003).

The analysis detailed herein focuses on three species: *Passer domesticus* (house sparrow), *Streptopelia decaocto* (collared dove) and *Accipiter nisus* (Eurasian sparrowhawk). This as over the past 40 years there has been an observed abundance decline in *P. domesticus* ($\sim 60\%$) (Robinson et al., 2005). One possible cause is that of an increase in the abundance and distribution of the *A. nisus*, over a similar time period (Newton, 1986). Besides this, *S.*

decaocto have colonised the UK in a similar time frame but in the reverse direction to *A. nisus* so act as a natural control species for testing potential effects of predators (Thomson et al., 1998; Newson et al., 2010; Swallow et al., 2016). Biodiversity monitoring focuses on specific indicator species such as songbirds—like *P. domesticus*—which are more accessible than other species and particularly vulnerable to changes in ecosystems due to their position high up in the food chain (Bibby et al., 2000; Gregory and van Strien, 2010). In order to fully appreciate, and gain a better understanding of the relationships between raptors and songbirds, it is necessary to analyse the different types of complex interactions operating among several species in space and time. Of particular interest may be inter-species interaction that may occur between a predator and its prey or between two sympatric species. In particular, one might want to infer about the relationship between species on different or the same trophic levels, respectively.

As mentioned above the data considered in this section are of a semi-continuous nature, and to account for this a delta gamma spatio-temporal model—a mixture distribution with probability p of observing a zero and $1-p$ of observing an observation from a gamma distribution (Aitchison, 1955)—is used to model each response variable. The section below outlines a multivariate spatio-temporal delta gamma model.

6.2.1 The Model

A Delta-Gamma model is used to represent the biomass of each avian species. Let z_{ik} be a binary indicator of the k^{th} species' presence ($k = (1, 2, 3)$ = (sparrowhawk, collared dove, house sparrow)) at site i. Then $z_{ik} \sim \text{Bernoulli}(p_{ik})$, where p_{ik} is the probability of presence of the k^{th} species at site i. Letting d_{ik} be the density of the k^{th} species at location i, then d_{ik} is given by,

$$d_{ik} = \begin{cases} \text{Gamma}(a_{ik}, b_{ik}) & \text{with probability } p_{ik} \\ 0, & \text{otherwise,} \end{cases} \tag{6.1}$$

with shape and scale parameters (a_{ik}, b_{ik}) respectively, and, $E[d_{ik}] = \mu_{ik}$.

Note, the binary components of the response reveal at which site each species is most likely to appear, whereas the non-zero component reveals the spatial distribution of the abundance for each species. This facilitates accounting for both the remaining spatial auto-correlation and dependence among the species. Utilising multiple latent stochastic structures in the modelling procedure leads to the linear predictors for each species being constructed as,

$$
\begin{aligned}
A.\ nisus &\begin{cases} \text{logit}(p_{i1}) = \mathbf{x_1}\{\mathbf{s}_i, t\} \\ \log(\mu_{i1}) = \beta_1 \mathbf{x_1}\{\mathbf{s}_i, t\} \end{cases} \\
S.\ decaocto &\begin{cases} \text{logit}(p_{i2}) = \mathbf{x_2}\{\mathbf{s}_i, t\} \\ \log(\mu_{i2}) = \beta_2 \mathbf{x_2}\{\mathbf{s}_i, t\} \end{cases} \\
P.\ domesticus &\begin{cases} \text{logit}(p_{i3}) = \alpha + \beta_{z3}\mathbf{x_1}\{\mathbf{s}_i, t\} + \gamma_{z3}\mathbf{x_2}\{\mathbf{s}_i, t\} + \mathbf{x_3}\{\mathbf{s}_i, t\} \\ \log(\mu_{i3}) = \alpha_y + \beta_{y3}\mathbf{x_1}\{\mathbf{s}_i, t\} + \gamma_{y3}\mathbf{x_2}\{\mathbf{s}_i, t\} + \beta_3 \mathbf{x_3}\{\mathbf{s}_i, t\}, \end{cases}
\end{aligned} \tag{6.2}
$$

where each $\mathbf{x_j}\{\mathbf{s}_i, t\}, j = (1, 2, 3)$, is a spatio-temporal random effect modelled by a GRF (as detailed in Chapter 4). Each $\alpha.$ is an intercept term. The parameters $\beta.$, and $\gamma.$, are scaling parameters to the spatio-temporal random fields of which they are coefficients. Each spatio-temporal random field has the following interpretation:

- $\mathbf{x_1}\{\mathbf{s}_i, t\}$: the spatially varying process referring to the probability of presence of *A. nisus*, which is also assumed to be a contributory factor to the spatially varying density of *A. nisus* over the time period. One would expect β_1 to be both positive and statistically significant.

- $\mathbf{x_2}\{\mathbf{s}_i, t\}$: the spatially varying process referring to the probability of presence of *S. decaocto*, which is also assumed to be a contributory factor to the spatially varying density of *S. decaocto* over the time period. One would expect β_2 to be both positive and statistically significant.

- $\mathbf{x_3}\{\mathbf{s}_i, t\}$: the spatially varying process referring to the probability of presence of *P. domesticus*—conditional on the spatial distribution of both *A. nisus* and *S. decaocto*—which is also assumed to be a contributory factor to the spatially varying density of *P. domesticus* over the time period. One would expect β_3 to be both positive and statistically significant. That is, $\mathbf{x_3}\{\mathbf{s}_i, t\}$ accounts for any remaining structure in the spatial distribution of *P. domesticus* once it has been ascertained whether either the prey species—*A. nisus*—or sympatric species—*S. decaocto*—are considered to have an effect on the spatial distribution of *P. domesticus*. Thus, if the spatial distribution of *P. domesticus* were independent of either that of *A. nisus* or *S. decaocto* $\mathbf{x_3}\{\mathbf{s}_i, t\}$ would account for the variation specific to the spatial distribution of *P. domesticus*.

6.2.2 Results

The model in Equation 6.2 was fitted using a SPDE model for the random effects in `R-INLA`. The Delauney triangulation constructed for the model, along with the sites at which the data were collected, is shown in Figure 6.1. As the modelling is implemented in a Bayesian framework (INLA) some thought has to be given to the sensitivity of the estimator to priors

Table 6.1: Posterior means, standard errors, and 95% credible intervals, for the parameters of the joint spatio-temporal model in Equation 6.2. The forward slash separates parameter estimates pertaining to each of the models fitted using the two sets of priors mentioned in the text: before, $N(-1, 10)$ priors on the interaction parameters β_{y3} and β_{z3}, and $N(0, 10)$ on all other interaction parameters, after, $N(0, 10)$ priors on all interaction parameters.

				Quantile of Order	
	Parameter	Mean	SE	2.5%	97.5%
A. nisus	ρ_1	0.98/0.981	0.004/0.004	0.97/0.972	0.987/0.988
	β_1	0.542/0.532	0.041/0.041	0.463/0.451	0.625/0.61
S. decaocto	ρ_2	0.963/0.962	0.006/0.006	0.95/0.949	0.975/0.973
	β_2	0.607/0.688	0.046/0.059	0.517/0.571	0.699/0.804
P. domesticus	α	-0.23/0.846	0.384/0.357	-0.984/0.145	0.524/1.547
	α_y	1.173/1.341	0.07/0.067	1.035/1.21	1.31/1.472
	ρ_3	0.968/0.964	0.007/0.008	0.952/0.946	0.981/0.977
	γ_{z3}	1.114/1.619	0.193/0.251	0.73/1.139	1.49/2.123
	γ_{y3}	0.494/0.577	0.068/0.079	0.358/0.425	0.624/0.735
	β_3	0.542/0.532	0.041/0.041	0.463/0.451	0.625/0.61
	β_{z3}	-0.93/-0.448	0.226/0.21	-1.397/-0.869	-0.51/-0.046
	β_{y3}	-0.15/-0.09	0.05/0.049	-0.249/-0.184	-0.053/0.007

set on the interaction parameters. This as the model is constructed such that those parameters are assumed to reflect the interaction between the spatio-temporal dynamics governing each species' distribution. Presented here are two models—of the same construction—where a different set of $(N(\mu, \sigma^2))$ priors are used for the interaction parameters. In particular, (i) $N(-1, 10)$ priors on the interaction parameters β_{y3} and β_{z3}, and $N(0, 10)$ on all other interaction parameters, and (ii) $N(0, 10)$ priors on all interaction parameters.

Table 6.1 shows the posterior means, standard errors, and 95% credible intervals, for the parameters of the joint spatio-temporal model in Equation 6.2. The forward slash separates parameter estimates pertaining to each of the models fitted using the two sets of priors mentioned in the text: before, $N(-1, 10)$ priors on the interaction parameters β_{y3} and β_{z3}, and $N(0, 10)$ on all other interaction parameters; after, $N(0, 10)$ priors on all interaction parameters. In both cases each of $\beta_j, (j = 1, 2, 3)$ are positively significant. This, as expected, indicates that the non-zero densities of each species and the presence of those species share the same spatial patterning respectively. The positive posterior means—in both cases—of both γ_{z3}, and γ_{y3} suggest that the non-zero densities of *P. domesticus* and the presence of *P. domesticus* share the same spatial patterning as the respective *S. decaocto* variables. Furthermore, the estimates in Table 6.1 indicate that all AR(1) temporal correlation coefficients—ρ.—are close to one. This indicates that there is strong short-term temporal dependence

between the occurrences of each species.

Of particular interest are the interaction parameters β_{z3} and β_{y3}. These parameters are considered to represent the interaction between the spatially varying probability of *A. nisus* presence and the probability of *P. domesticus* presence, and the spatial non-zero density of *P. domesticus* respectively. In each case considered, the negative posterior mean of β_{z3} suggests that the spatial pattern formed by the presence of *A. nisus* is negatively related to the *P. domesticus* probability of presence (i.e., sites that are more likely to be occupied by *A. nisus* are less likely to be occupied by *P. domesticus*). It should be noted that the 95% credible interval for the parameter β_{y3} (just) contains zero in the case when a $N(0, 10)$ is given, and does not when given a $N(-1, 10)$ prior (Table 6.1). The posterior mean for β_{y3} however is negative in each case, thus perhaps suggesting that there may also exits a negative interaction between the spatially varying probability of *A. nisus* presence and spatial non-zero density of *P. domesticus*.

One other point to note is the estimates for the intercept parameter α, essentially the baseline inverse logit probability of observing *P. domesticus* (ignoring any spatial variation). In each model the standard errors of these estimates are large, reflecting the uncertainty regarding the estimate. Now in the case when the non-zero density of *P. domesticus* is estimated to be negatively related to the probability of *A. nisus* presence α could conceivable be zero. Whereas in the case when the non-zero density of *P. domesticus* is estimated not to be negatively related to the probability of *A. nisus* presence α (on the response scale) is $\simeq 0.7$ (with standard error 0.357), which indicates that if *P. domesticus* were not adversely affected by the presence of *A. nisus* then (on average) the probability of observing them would be higher.

Figure 6.2 shows—for both fitted models—the mean contribution of each species specific random effect to the probability of presence of *P. domesticus* over the time period. Line colours refer to the models fitted using the two sets of priors mentioned above: black, $N(0, 10)$ priors on all interaction parameters, grey, $N(-1, 10)$ priors on the interaction parameters β_{y3} and β_{z3}, and $N(0, 10)$ on all other interaction parameters. The dotted lines illustrate the average contribution of the process governing the spatial distribution of *S. decaocto* to the probability of presence of *P. domesticus* ($\gamma_{z3}\,\mathbf{x_2}\{\mathbf{s}_i, t\}$). Dashed lines illustrate the average contribution of the process governing the spatial distribution of *A. nisus* to the probability of presence of *P. domesticus* ($\beta_{z3}\,\mathbf{x_1}\{\mathbf{s}_i, t\}$). The dot-dash lines represent the average of the *P. domesticus* specific random effect in each year ($\mathbf{x_3}\{\mathbf{s}_i, t\}$).

As illustrated by Figure 6.2, both $\beta_{z3}\,\mathbf{x_1}\{\mathbf{s}_i, t\}$, and $\mathbf{x_3}\{\mathbf{s}_i, t\}$ decrease the probability of presence of *P. domesticus*. Moreover, $\gamma_{z3}\,\mathbf{x_2}\{\mathbf{s}_i, t\}$ has a positive effect on the probability of presence of *P. domesticus*. That is, the presence of *P. domesticus*, do decrease over time—reflecting their observed demise—yet, processes promoting the presence of *A. nisus* decrease the presence (and possible density) of *P. domesticus*. The processes promoting the presence

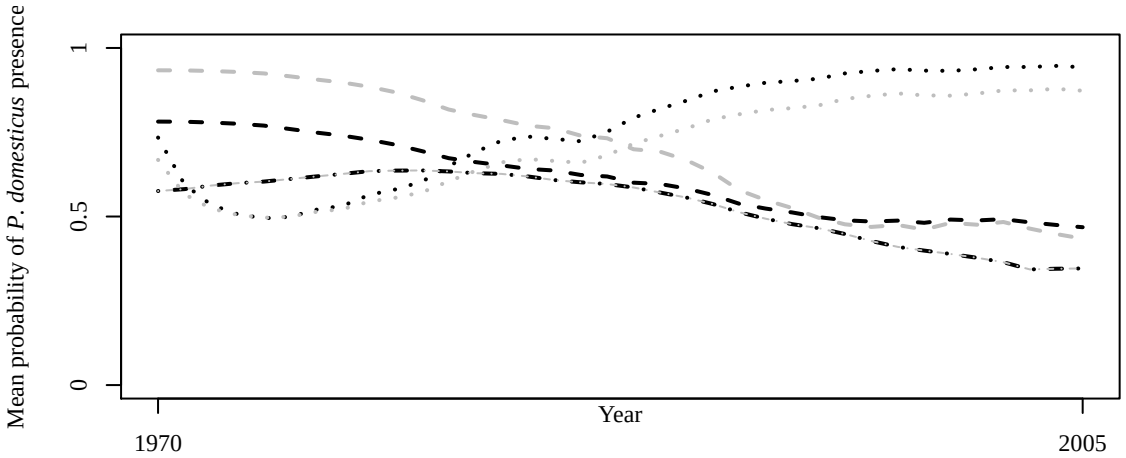


Figure 6.2: Plot showing the mean contribution of each species specific random effect to the probability of presence of *P. domesticus* over the time period. Dotted lines–the average contribution of the process governing the spatial distribution of *S. decaocto* to the probability of presence of *P. domesticus*. Dashed lines–the average contribution of the process governing the spatial distribution of *A. nisus* to the probability of presence of *P. domesticus*. Dot-dash lines–the average of the *P. domesticus* specific random effect in each year. Line colours refer to the models fitted using the two sets of priors mentioned in the text: black, $N(0, 10)$ priors on all interaction parameters, grey, $N(-1, 10)$ priors on the interaction parameters β_{y3} and β_{z3}, and $N(0, 10)$ on all other interaction parameters.

of *S. decaocto* are in accordance with the probability of *P. domesticus* presence. Thus, as expected the results indicate a sympatric relationship between *P. domesticus* and *S. decaocto*, and a predator-prey relationship between *P. domesticus* and *A. nisus*.

This section utilised shared stochastic structures to infer multi-species relationships. The mathematical construct of the class of model discussed is extremely flexible. In particular, the model construction may be thought to reflect the dependency structure inherent in many ecological datasets. As such, the following section utilises similar methodology to model other dependency structure which operates in the context of ecology.

6.3 A predictive habitat dynamics model

The section above focuses on inferring multi-species interactions; yet, many other varieties of ecological relationships exist. For example, abiotic relationships, which are interactions between physical parts of the environment and a species. Consider—in the context of criminology—house burglary, such crimes are clearly dependent on the locations of houses. Yet, there exist mechanisms—something akin to the process governing burglar's choice—which are conditional on there being a house at that location. Specifically, conditional on there being a house at some considered location certain other aspects of the environment may increase the likelihood that a particular house is targeted. Along a similar vein, intra-species distribution is conditional on habitat. That is, given that a species has a preference for a certain type of habitat, there are extraneous factors which affect the spatial distribution of that species.

Accounting for such interactions may aid in understanding other unconsidered biotic or abiotic relationships, or indeed the data collection/observation process. As such, this section utilises some of the concepts introduced in Chapter 5 to infer the spatial distribution of a species conditional on its habitat dynamics. Incorporating habitat dynamics in a species distribution model is fundamental in order that the existing dependence is captured. As such, many ecological statistical methods aim at discovering the interaction between habitat and the considered species. This, as typically species prefer certain features within a habitat, due to either one of, or both biotic, and abiotic components. In addition to this, there is interest in the spatial dynamics of the considered species, conditional on the dependence to the structure of the habitat. Therefore, in order to infer the spatial intra species dependence, the spatial structure of the habitat must be accounted for.

Between 1994–2004 the population of *A. noctua* has seen a decrease of $\sim 24\%$ in Germany, and $\sim 50\%$ in Poland (Van Nieuwenhuyse et al., 2008). In this chapter *A. noctua*s abundance in the Czech Republic is considered due to decreasing populations in the north of Slovakia. This showing that *A. noctua*s are, in recent years, also under pressure (Šálek and Schröpfer, 2008). In certain parts of Slovakia, including the region discussed herein, it has been well established that since 2010 *A. noctua* has solely resided in farm buildings (Dobrỳ, 2011; Šipkovskỳ, 2012).

The data considered comprise of the locations of farms in the South West—$47.75°N$, $17.07°E$ - $48.89°N$, $18.96°E$—region of Slovakia, some of which were visited to ascertain if any *A. noctua* were present (see Figure 6.3). Data collection was carried out from just after sunset to approximately eleven at night. Initially, recordings of calls—of *A. noctua*—were played for three minutes followed by two minutes of listening. Then recordings were played for a further two minutes. If no responses were heard—from *A. noctua*—after a further five minutes of listening, the area was considered unoccupied (see Dobrỳ (2011) for more details). Perfect detection from the survey methodology is assumed. The modelling framework proposed herein enables one, conditional on farm locations, to infer the spatial structure of the presence/absence of *A. noctua*).

As in most cases of census data collection, attempting to encounter all presences of *A. noctua* would be impractical, mainly due to prohibitive financial and time constraints. However, the geolocations of farms are readily available[2]. Thus, incorporating that information into the methodology developed herein enables one to predict at which unvisited farm *A. noctua* is most likely to reside. This methodology's main focus is inferring the spatial distribution of *A. noctua* given all possible inhabitant locations. In addition, the proposed framework facilitates incorporation of a misaligned covariate—that is a covariate not observed at the point pattern locations. Typically, covariates also have spatial dependence, thus the model has to enable prediction of the covariate at the data collection locations (a more in-depth

[2]from sources such as the `https://developers.google.com/places/` API

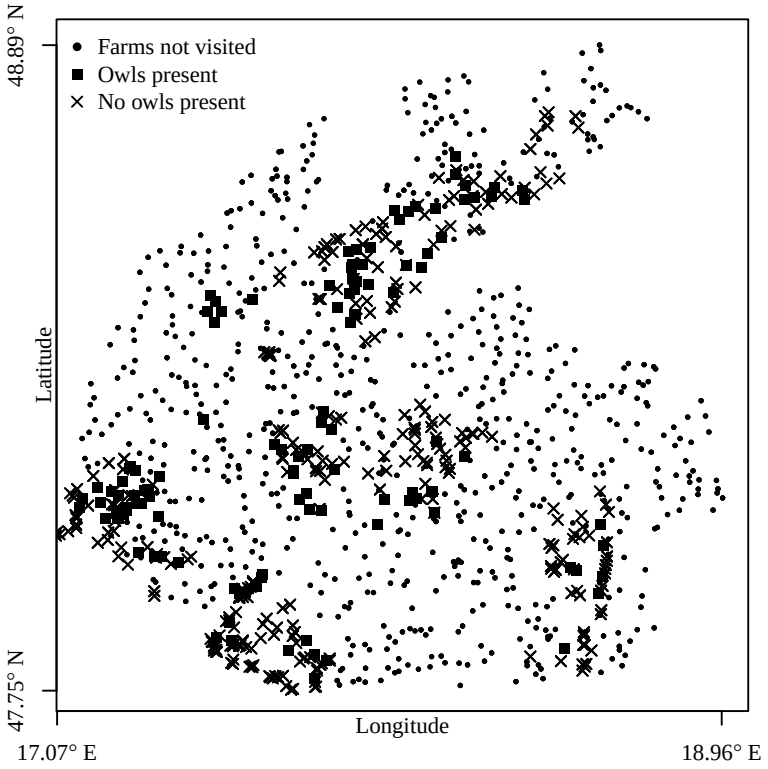


Figure 6.3: Plot of farm locations in the South-West region of Slovakia. Crosses indicate the farms visited where no *A. noctua*s were found, whereas the filled squares indicate farms visited where *A. noctua*s were detected. The filled circles refer to farms not visited, yet were incorporated into the model in order to use the spatial information of these "possible" habitat location.

discussion of this can be seen in Blangiardo and Cameletti (2015, Chapter 8)). Letting $Y(\mathbf{s})$ be the binary presence/absence of *A. noctua* at each farm, a farm location not visited is given as NA, this enables prediction at these locations. Then, $Y(\mathbf{s}) \sim$ Bernoulli($p(\mathbf{s})$ where $p(\cdot)$ is the probability of the presence of *A. noctua*. In addition, letting $w(\mathbf{s}) \sim$ Gamma$(a(\mathbf{s}), b(\mathbf{s}))$ be the observed snowfall in centimetres (observed at different locations than the farms) where the expected snowfall is given by $\mathbf{z}_2(\mathbf{s}) = \frac{a(\mathbf{s})}{b(\mathbf{s})}$. Then to incorporate this misaligned covariate into the modelling framework a joint model of the form given below is constructed.

$$\begin{aligned} \text{logit}(Y(\mathbf{s})) &= \boldsymbol{\beta}\,\mathbf{x} + \mathbf{z}_1(\mathbf{s}) + \gamma\,\mathbf{z}_2(\mathbf{s}), \\ \log(w(\mathbf{s})) &= \mathbf{z}_2(\mathbf{s}). \end{aligned} \tag{6.3}$$

Here $\mathbf{z}_2(\mathbf{s})$ is the spatial effect reflecting average snowfall shared between likelihoods, its contribution to the spatial distribution of *A. noctua* is encompassed by the interaction parameter γ. The random effect $\mathbf{z}_1(\mathbf{s})$ reflects the spatially varying distribution of *A. noctua* conditional on farm locations and snowfall. Covariates collected at the farm locations are given by $\mathbf{x}$ with coefficients $\boldsymbol{\beta}$. This analysis considered the factor variable land type (arable, artificial, and orchard) as a fixed effect.

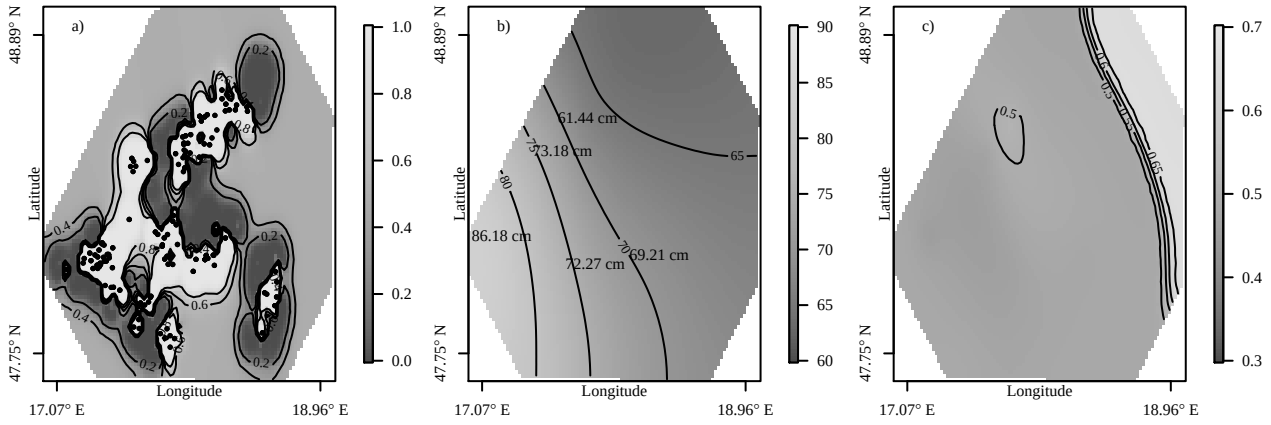


Figure 6.4: Plots showing the posterior mean of each GRF in Equation (6.3). a) shows the estimated mean of $\mathbf{z}_1(\mathbf{s})$ (with farms locations where *A. noctua* were present), b) the estimated mean of $\mathbf{z}_2(\mathbf{s})$ (with the observed snowfall given at the location of each meteorological station), and c) $\gamma\,\mathbf{z}_2(\mathbf{s})$. Each field is shown on its respective response scale (i.e., $\mathbf{z}_1(\mathbf{s})$ and $\gamma\,\mathbf{z}_2(\mathbf{s})$ on the probability scale, and $\mathbf{z}_2(\mathbf{s})$ reflecting the average snowfall in centimetres).

6.3.1 Inference

The joint modelling framework proposed enables on to infer the effect of average snowfall, in centimetres, on *A. noctua* as well as the process governing the (spatial) probability of presence of *A. noctua* conditional on habitat. This structure is shown in isolation by plot a) in Figure 6.4 (i.e., $\mathbf{z_1(s)}$). Recall that average snowfall(cm) was not recorded at the response locations, but obtained from the meteorological stations in the area (see Plot b) Figure 6.4). Thus, this misaligned covariate is modelled to infer its spatial structure, $\mathbf{z_2(s)}$, where the interaction parameter γ captures the relationship between this and the probability of presence of *A. noctua*. Plot c) Figure 6.4 shows, in isolation, the contribution of the shared GRF $\mathbf{z}_2(\mathbf{s})$ to the probability of presence of *A. noctua*; this is the probability of observing *A. noctua* if there were no contribution from the spatial effect represented by $\mathbf{z_1(s)}$. From comparing Plots b) and c) (Figure 6.4) an adverse relationship between snowfall and the probability of presence of *A. noctua* is inferred. That is, snowfall (broadly) decreases from South-West to the North-East, which is estimated to have an adverse effect on the probability of presence of *A. noctua*. It should be noted that the smoothness of these representative plots is down to the lack of information regarding snowfall included in the model (only information form 5 meteorological stations was available and used.

In addition to a misaligned covariate, information regarding the nearby land type at each farm location was collected and included as a fixed effect in the model. This variable was aggregated to be in one of three categories: arable, artificial, and orchard. The model estimated no discernible difference between these land types in predicting the probability of presence of *A. noctua*. The estimated coefficients (on the response scale) relating to each land type were 1.24×10^{-5}, 1.24×10^{-5}, and 1.34×10^{-5} respectively. This indicating that conditional on the farm locations *A. noctua* were slightly more likely to be found in orchards

than in areas of arable or artificial land.

This section illustrates that through conditioning on farm locations the spatial probability of presence of *A. noctua* can be inferred (Plot a) Figure 6.4), conditional on the structure of their habitat. In addition, the modelling framework detailed herein enables the mark—probability of presence of *A. noctua*—to be predicted at each non-visited farm. Thus, the spatial structure of the process governing the probability of presence can be inferred. Figure 6.4 illustrates that the model predicts (based on the observed data) a higher chance of presence of *A. noctua* towards the South-West of the considered region.

6.4 Discussion

This chapter extends methodology introduced in Chapter 4, in a similar vein to the methodology discussed in Chapter 5. Specifically, through the use of Bayesian methodology—Laplace Approximations (Rue et al., 2009)—coupled with using a SPDE (Lindgren et al., 2011) to model the assumed latent structures in the model. The chief focus of this chapter has been on using the methodology detailed to estimate interactions inherent in the ecological data discussed. In particular inferring,(i) relationships between avian species on the same and different trophic levels, and (ii) conditional on habitat the spatial dynamics of *A. noctua* in Slovakia.

The model discussed in Section 6.2 assess the spatio-temporal structure inherent in zero-inflated data relating to counts of *P. domesticus* obtained from the GBFS over 36 years. Both the relationship between sympatric and predator-prey species are accounted for in a joint model of several species. Multi-species interactions are highly relevant within ecology particularly when referring to the change in spatial distribution of species over time. The methodology introduced in Section 6.2 directly accounts for the spatial and temporal correlation, considering multiple avian species, over large regions, and over long time periods. Upon application of this methodology to the multi-species data detailed in Section 6.2 a significant negative correlation between the probability of sparrowhawk presence and the abundance of house sparrows was detected. This suggests that if a causal relationship exists, then sparrowhawks may be reducing the overall abundance of house sparrows observed, but have no significant effect on the probability of a house sparrow being observed.

It should be noted that not only can the methodology developed herein be applied to geostatistical data (i.e., spatially continuous measurements taken at a finite number of locations, as detailed in Section 1.1), but, where the interest may be to analyse the spatial pattern formed by objects or events in space. Or, in particular, where the objective is not to infer the intensity measure of the considered point pattern, but, to infer the mark process conditional on the point locations of the observed marked point pattern. Specifically inferring abiotic as opposed to biotic relationships.

Section 6.3 details an application to data pertaining to *A. noctua*; here, given their habitat structure—locations of farm buildings—the process relating to the spatial dynamics of the probability of their presence is inferred. Furthermore, two additional complexities (perhaps most) relevant to the field of ecology, are accounted for during the modelling procedure. These being (i) a covariate not sampled at the response locations (misaligned), and (ii) surveying—due to feasibility—only carried out at certain locations. This application utilises multiple GRFs, akin to Section 6.2, to infer the spatial distribution of *A. noctua* conditional on all possible inhabitant locations. A negative correlation between estimated average snowfall(cm) was detected. In addition, areas (not surveyed) South-West of Trnava $(48.3709°N, 17.5833°E)$ and East of Bratislava $(48.1486°N, 17.1077°E)$ were estimated to be areas most likely to be occupied by *A. noctua*.

This chapter chiefly focuses on incorporating the modelling of both abiotic and biotic relationships inherent in ecological data. Obviously there exist more complex dependency structures; yet, mainly due to computational and data restrictions, no one modelling framework could infer all the processes of interest. However, utilising the methodology detailed herein—that is specifically employing stochastic structures in the model to soak up the unexplained variation pertaining to variables—enables one to assume that some structure contributes to the observed structure of multiple variables. In particular, these stochastic structures can be thought of as representatives of the processes inherent in either abiotic or biotic relationships. In a continuation of this, this discussion conjectures (below) how this approach could be generalised to incorporate species interaction and habitat dynamics. Expressly illustrating the concept through a dataset pertaining to the population dynamics of ants (*Azteca sericeasur*) considered by Jackson et al. (2014). The remainder of this discussion conceptualises generalising the methodology developed in this chapter in fitting a marked spatio-temporal point process model to capture species interaction conditional on habitat dynamics in space and over time.

6.4.1 Modelling habitat dynamics: a point process model with two marks

In an extension to the analysis detailed in Section 6.3 the methodology developed herein may be employed to consider further species population dynamics, and environmental interaction. Data analysed by Jackson et al. (2014) pertains to the population dynamics of a tree nesting ant, *A. sericeasur*. Here, the data were collected on a 300 hectare coffee plantation situated in the Socousco region of Chiapas, Mexico (15.11° N, 92.20° W), 2005-2011. Locations of all trees in the plantation were recorded, and it was noted whether they contained a *A. sericeasur* nest; also it was recorded if there existed a nearby coffee bush infested by *Coccus viridis* coffee green scale. It is widely known that there exists a mutualistic relationship between *A. sericeasur* and *C. viridis*; *A. sericeasur* cultivate *C. viridis* on nearby coffee bushes harvesting a sugar-rich honeydew excreted by the scales. In addition, *A. sericeasur*

safeguard the *C. viridis* through attacking their natural enemies. Jackson et al. (2014) found an increase in the population of the *A. sericeasur*, despite a reduction in the number of potential nesting sites (approximately 30% of trees being felled over a two year period, 2007-2008). Moreover, the *A. sericeasur* was found to maintain "spatial self organisation" throughout the period (although the local clustering of the nests was reduced post-pruning). Figure 6.5 illustrates the structure of the data, which was simulated to reflect the data detailed by Jackson et al. (2014). Two realisations of the spatio-temporal marked point process are shown—conjectured to refer to the years 2005, and 2011 (pre and post felling respectively). The black circle plotting characters in Figure 6.5 refer to trees inhabited by *A. sericeasur*; the grey encircled characters refer to trees with nearby *C. viridis* infested coffee bushes. The positive spatial relationships between *A. sericeasur* and *C. viridis*, simulated to reflect their mutualistic relationship, is evident in both plots. Moreover, there is a clear decrease in the number of trees—reflecting the large scale felling which occurred over a two-year period—and a seeming increase in the proportion of trees inhabited by the *A. sericeasur*. In addition, local scale clustering of *A. sericeasur* inhabited trees is evident in both plots, yet there is a marked decrease in the degree of spatial clustering in 2011 (post felling of trees).

Now, one may consider the (homogeneous) tree locations a realisation of a marked point process, the marks being:(i) a binary variable indicating if the tree is occupied by a *A. sericeasur* nest, and (ii) a binary variable indicating if there is a nearby coffee bush infested by the pest *C. viridis* . The mathematical construction of such a model is similar to that discussed in Section 6.3. Letting the tree locations be a realisation of a LGCP, $N(\Omega)$, and the random variables A and S, pertain to each mark respectively, then

$$\begin{aligned} N(\Omega) \sim \text{Poisson}(\Lambda(\mathbf{s},t)) &\rightarrow E[N] = \exp(z(\mathbf{s},t))) \\ A_i \sim \text{Bernoulli}(p_i) &\rightarrow E[A] = \text{logit}^{-1}\left(z(\mathbf{s},t) + z_1(\mathbf{s},t)\right) \\ S_i \sim \text{Bernoulli}(b_i) &\rightarrow E[S] = \text{logit}^{-1}\left(z(\mathbf{s},t) + \beta\, z_1(\mathbf{s},t)\right) \end{aligned}$$

where p_i and b_i are the respective probabilities that a tree contains an *A. sericeasur* nest or has a *C. viridis* infested coffee bush nearby. Each $z.(\mathbf{s},t)$ is a spatio-temporal random field, where the interaction parameter β captures the dependence between the spatial distribution of *A. sericeasur* nests and the *C. viridis* infested coffee bushes conditional on the loactions of trees. Thus, utilizing multiple latent GRFs the spatio-temporal dynamics of *A. sericeasur* population and the (mutualistic) relationship with the *C. viridis* insect can be captured. It should be noted that the data is simulated from the above model, as an illustrative notion only—no model fitting is discussed here.

Applying such a modelling approach to the data discussed above, would, one would conjecture, result in estimation of the spatial distribution of trees. How this changes over the time period is clearly not of any real interest (as farmyards in *A. noctua* data above). However, conditional on the spatial distribution of trees (the habitat of the *A. sericeasur*) the structure $z_1(\mathbf{s},t)$ would reflect the process describing the spatially (and temporally) varying

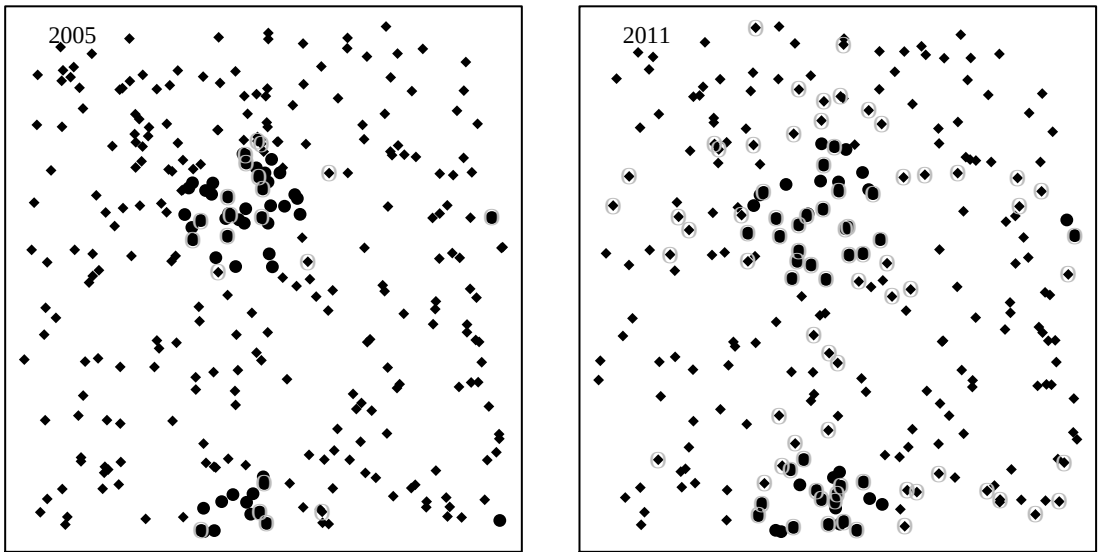


Figure 6.5: Simulated data reflecting the structure of the data detailed by Jackson et al. (2014) show in years 2005—prior felling—and 2011—post felling. Locations of all trees are plotted, circle plotting characters refer to trees inhabited by *A. sericeasur*s, and the grey encircled characters refer to trees with nearby *C. viridis* infested coffee bushes. From these plots it is clear the positive spatial relationships between *A. sericeasur*s and *C. viridis*s, simulated to reflect their mutualistic relationship. Moreover, there is a clear decrease in the number of trees—reflecting the large scale felling which occurred over a two-year period—and a seeming increase in the proportion of trees inhabited by the *A. sericeasur*. In addition, local scale clustering of *A. sericeasur* inhabited trees is evident in both plots, yet there is a marked decrease in the degree of spatial clustering in 2011 (post felling of trees).

probability of an area being inhabited by *A. sericeasur*s. Thus, despite the drastic change in the number of trees (due to a two year period of felling) $z_1(\mathbf{s}, t)$ would capture the (change in) spatial dymanics of *A. sericeasur* colonies in the study region. In addition, the same procedure enables the spatial relationship between the *A. sericeasur* and *C. viridis* insect to be estimated over time.

The use of multiple latent GRFs in modelling complex spatial, and spatio-temporal dynamics is inherent in many fields of study has been a core concept in Chapters 4 and 5 as well as in this chapter. Specifically, estimation of the interaction parameter—herein typically denoted $\beta_\cdot$ or $\gamma_\cdot$—the coefficient of some GRF. These stochastic structures can be thought of as reflecting the first order dynamics of the observations: reflecting characteristics of the underlying distribution. However, in many instances it is the second order behaviour that is of interest. Second order dynamics typically refer to second level characteristics. These being characteristics not purely operating over the observed data, but describing how the characteristics of the underlying distribution behave. Consider a LGCP, the first order characteristics of this measure would be its intensity measure, a second order characteristic might be a process governing the variance of this intensity measure (Note: **not** the variance of the LGCP). The following chapter of this thesis chiefly focuses on exploring the second order characteristics of LGCPs, specifically through considering non-stationary LGCPs (full details of the meaning of non-stationarity are given in the chapter).

Part V

Second order structures of latent Gaussian random fields

Chapter 7

Non-stationary log-Gaussian Cox processes

7.1 Introduction

Chapters 4, 5, and 6 detailed the use of latent Gaussian models–primarily in the context of fitting LGCPs—to account for spatial dependence inherent in spatially indexed data. These data ($\mathbf{y}$) are considered to be independent conditional on a latent field ($\mathbf{z}$) with parameters, $\boldsymbol{\theta}$. That is,

$$\mathbf{y}|\mathbf{z}, \boldsymbol{\theta} \sim \pi(\mathbf{y}|\mathbf{z}, \boldsymbol{\theta}) = \prod_{i=1}^{N} \pi(y_i|z_i, \boldsymbol{\theta}).$$

Here the latent field $\mathbf{z}$ is Normally distributed, $N(\mathbf{z}; \boldsymbol{\mu}(\cdot), \mathbf{Q}^{-1}(\cdot))$, with mean $\boldsymbol{\mu}$ (in the case of the GMRFs discussed here $\boldsymbol{\mu} = \mathbf{0}$) and precision $\mathbf{Q}$. The methodology discussed in those chapters modelled the latent field $\mathbf{z}$ using a SPDE model (Lindgren et al., 2011). This through considering the SPDE given by,

$$(\kappa^2 - \Delta)\, \mathbf{z}(\mathbf{s}) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^d \quad \kappa \geq 0, \tag{7.1}$$

the solution to which is—under conditions detailed below—the mean-zero Gaussian random field $\mathbf{z}(\mathbf{s})$. In Equation (7.1) above , $\Delta = \sum_{i=1}^{\mathrm{d}} \frac{d^2}{ds_i^2}$ denotes the Laplacian, κ is the spatial scale parameter, and $\mathcal{W}(\mathbf{s})$, is a Gaussian spatial white noise process. Equation (7.1) is equivalent to the SPDE given in Equation (1.5) where the parameter α has been fixed at 2, and τ set to 1.

As detailed in Section 1.3.2 Lindgren (2012) derive the precision matrix of $\mathbf{z}$ as,

$$\mathbf{Q} = \kappa^4 \mathbf{C} + 2\kappa^2 \mathbf{G}_1 + \mathbf{G}_2, \tag{7.2}$$

where $\mathbf{C}$, $\mathbf{G}_1$, and $\mathbf{G}_2$ are sparse matrices obtained from the finite basis representation (Equation (1.7)). The parameter κ, the spatial scaling parameter, is related to the correlation range

of the field. That is, κ describes the extent of the spatial dependence in the field. In particular Lindgren et al. (2011) quote the empirically derived distance at which the correlation (in d dimensions) in the field drops to 0.1 to be,

$$\rho = \frac{\sqrt{8\,\nu}}{\kappa}, \quad \text{where} \quad \nu = \alpha - \frac{d}{2}.$$

This range parameter, ρ, can be thought of as the rate of decay, with larger values of ρ (smaller κ) corresponding to more highly correlated observations. From hereon the spatial dependence of the field is discussed through referring to the parameter κ.

The underlying assumptions of the solution to the SPDE given by Equation (7.1) above having precision given in Equation (7.2) are:(i) **isotropy**, no directionality in the field—the precision is a function of the Euclidean distance only, and (ii) **stationarity**, the local stochastic structure of the field is similar across the domain—the mean is constant and the precision is a function of the difference in (spatial) location only. The majority of point process models to date assume stationarity (Illian et al., 2013, 2012a; Barthelmé et al., 2013). Specifically, through assuming that observing a certain configuration at a particular location is equally likely independent of the location. There exists no all-encompassing statistical test that proves that a point pattern is a realisation of a non-stationary process, as, in practice, only a single realisation of a point process is typically observed. Thus, when no comparisons can be made it is difficult to tell apart second order from first order stationarity. Therefore for either practical reasons, or simplicity, stationarity is commonly assumed. Justification for not assuming stationarity is purely based on either(i) information gleaned from the data, or (ii) through scientific argument.To propose non-stationarity based on the data's structure alone is sensitive to the spatial scale at which the point pattern is observed. This as, differentiating between the intra cluster pattern versus the inter cluster scale is impossible.

Consider locations of houses at which a house break in is reported. These locations may be thought of as a realisation of a LGCP. Modelling these data as a stationary process would assume that the point density is constant across the area. In addition, it may be assumed that the local point configurations (point dependence) does not change with location. However, burglaries may(i) be systematically driven by some environmental variable (i.e., more likely to occur in higher populated areas), (ii) display different spatial configurations dependent on spatial location (i.e., sparse clusters in a rural region as opposed to urban areas). The examples given are of course exceptionally simplistic, and may be accounted for through some environment based spatial fixed effect.

On the other hand, however, it may not be the local environment that leads to non-stationarity of the observed point pattern, rather the nature of the data themselves. Consider introduction of a species to a region, here it would not be the environment that affected the distribution of that species, but time. The introduced species will initially only be able to react to their local environment(i.e., that close to their introduction site); full colonisation of

the environment will not have occurred.

In the context of LGCPs this means that the cluster configuration is dependent on location non-systematically, and typically not relatable to some known environmental factor. Thus, the second order behaviour of the process is also stochastic. That is, the spatial dependency structure of the assumed latent GRF is not constant in space. Likewise, the configurations of clusters may be random.

In many cases, the stationary assumption may not be justifiable: the spatial dependence between observations, and/or, the mean structure may vary across the domain. Consider an isotropic GRF, $\mathbf{x_s}$, such that $E[\mathbf{x_s}] = 0$ and $Cov(x_i, x_j) = \mathcal{C}$, then two conceivable ways in which non-stationarity might be reflected are as follows:

1. Letting $\mathbf{z_s} = \mathbf{m_s} + \mathbf{x_s}$ where $\mathbf{m_s}$ is a real valued function. Then $E[\mathbf{z_s}] = \mathbf{m_s}$ and $Cov(z_i, z_j) = \mathcal{C}$, (i.e., non-stationary expectation). For example, the mean structure of the observations changes in space. Consider species abundance, then a non-stationary mean structure assumes that average abundance changes across space, possibly due to a species-environment relationship.

2. Letting $\mathbf{z_s} = \sigma(\mathbf{s})\,\mathbf{x_s}$, where $\sigma(\mathbf{s})$ is a real valued function. Then $E[\mathbf{z_s}] = 0$ and $Cov(z_i, z_j) = \sigma_i\,\sigma_j\,\mathcal{C}$, (i.e., non-stationary variance specified by $\sigma^2(\mathbf{s})$). Here, the spatial dependence between observations varies across space. This second order behaviour might reflect more similarity in observations in certain areas, conversely in others there may be no such dependence.

The former case, although computationally expensive, is a relatively simple extension of the LGCP model discussed in Chapters 4, 5, and 6, and is not discussed here. In the latter case, there is a distinction to be made between(i) a precision structure that changes with respect to some known (fixed) effect, that is, some real valued function which changes in space; and (ii) a precision structure that changes in space with respect to some unknown effect. The former is akin to fitting a linear regression to some parameter of the precision. The latter is akin to assuming a stochastic structure on a parameter of the latent field. From herein, non-stationary models refer only to non-stationarity in the precision of the GRF.

The difference between systematic and second order non-stationarity is illustrated in Figure 7.1. Both non-stationary expectation and variance are illustrated. Specifically, from the left hand side of the plot to the right(i) point density decreases systematically (non-stationary expectation), and (ii) sparsity of the clusters increases (non-stationary spatial dependence).

Non-stationary models have been proposed and used in the fitting of models to geostatistical data (Lindgren, 2012; Ingebrigtsen et al., 2014). These assume the spatial dependence of the GRF changes with respect to some real valued function of space. Specifically, a linear regression is fitted to some function of the precision matrix. In particular, Ingebrigtsen

et al. (2014) assume that the spatial dependence structure of the random field in their model changes with respect to the local topography—gradient and elevation—of the area. This methodology is hindered by the requirement of some environmental fixed effect in line with which the spatial dependence is assumed to change.

In addition to non-statationarity, another commonly discussed second order behaviour of GRFs is oscillation. That is, some repetitive variation of the precision structure over space. Specifically $\mathbf{Q}$ (Equation (7.1)) varies systematically—oscillates. The subtle difference between oscillation and non-stationarity should be noted. Specifically, both refer to spatially varying precision, yet the oscillatory behaviour is irrespective of the spatial domain. In particular, oscillatory behaviour is typically parameterised using either sine or cosine functions, not functions of spatial location. As such, only non-stationary models—spatially varying κ (Equation (7.2))—result in a spatially varying marginal variance.

Estimating the second order behaviour requires methodology which is both flexible and capable of dealing with the computational burden in reasonable time. As such, typically approximation methods are employed over simulation based approaches. In particular, when assuming a SPDE model for the latent field the authors of `R-INLA` Rue and Martino (2007) have implemented incorporation of non-stationarity (through a fixed effect) in their package, see Lindgren (2012); Ingebrigtsen et al. (2014); Blangiardo and Cameletti (2015). In recent years other software has been developed named Template Model Builder (TMB) (Kristensen et al., 2016). TMB is based on AD and Laplace approximation techniques (Fournier et al., 2012; Albertsen et al., 2015; Kristensen et al., 2016). Due to the user being able to specify any model type some consider TMB to be more flexible than INLA. It should be noted however that the model types available in INLA are extensive and the typical user is unlikely to require any specification beyond those readily available. In addition, TMB does require the user to specify specific template types based on C++ syntax, which users may not find as user friendly as INLA.

Both approaches—INLA and TMB—are specifically designed to speed up the computational time in the fitting of random effect models. The estimator TMB is based on the AD Model Builder (ADMB) package (Fournier et al., 2012). Considering the joint likelihood for the data and the random effects, TMB maximises the Laplace approximation of the marginal likelihood automatically integrating out the random effects. The approximation, and derivatives, are obtained using AD of the joint likelihood. AD is a technique, given some computer algorithm defining a function, used to numerically compute the derivatives of that function. It is beyond the scope of this thesis to discuss AD in any more detail, yet if of interest the reader should see Griewank and Walther (2008); Fournier et al. (2012). INLA uses Laplace approximations as detailed in Section 1.3.1. In summary, the joint posterior marginal of the parameters $\boldsymbol{\theta}$ are approximated by Laplace's method, this Gaussian approximation is then improved by again using Laplace's method (or the simplified Laplace approximation) for se-

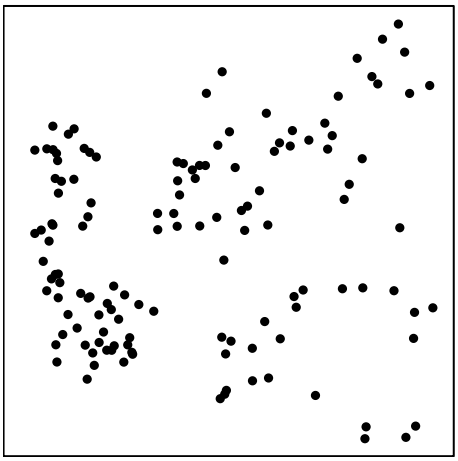


Figure 7.1: An illustration of a non-stationary cluster point pattern. Here both the expectation and the variance are non-stationary. The spatial configuration of the clusters is location dependent. The mean structure (point density) of the process decreases systematically from left to right, the sparsity of the clusters increases from left to right.

lected θ. In addition to the Laplace approximations, sparse matrix algorithms are exploited to further aid in the computational efficiency of the methodology. Both methodologies, based on Laplace approximations, and further algorithmic techniques (automatic differentiation or sparse matrix calculations), are alternatives to the traditional simulation based Markov Chain Monte Carlo (MCMC) approach. The sparsity of the precision matrix $\mathbf{Q}$ is an integral part of the computational efficiency of either methodology.

The remainder of the chapter presents methodology that models varying second order behaviour of the assumed GRF. In particular, through detailing methodology that uses both `R-INLA` and TMB, both of which utilise Laplace approximation, an integral approximation technique. In summary, considering some twice differentiable function $f(\mathbf{x})$, Laplace's method finds the maximum of $f(\mathbf{x})$ and applies a second order Taylor series approximation to its logarithm. When considering posterior distributions, the maximum is considered the mode of the distribution and the second order Taylor series corresponds to a Gaussian distribution. This leads to integrals which can be computed analytically, speeding up computation. It should be noted however that Laplace approximations (essentially Gaussian approximations) assume that the probability mass is represented by a local maximum, and in situations where this is not the case the technique fails.

The following section details methods for modelling both oscillatory precision (Section 7.2.1), and non-stationary spatial dependence (Section 7.2.2). One might consider incorporating both forms of these second order behaviours into one model, however, this is not considered herein.

7.2 Methodology

The LGCP fitted here are assumed to be isotropic. Isotropy may be thought of as comparable in some respects to stationarity. An isotropic point process is one where observing a certain point configuration at a particular location is equally likely independent of the rotation—recall that a stationary process this is independent of location. Stationarity, and isotropy are not equivalent however. A process can be non-stationary and isotropic or vice versa. Anisotropy essentially signifies directionality in a process (e.g., wind direction affecting fertiliser spraying). The distinction is made here to note that this section considers non-stationarity alone. Despite the distinctions made above between stationarity and isotropy due to the parallels between these characteristics non-stationarity is considered alone.

This chapter develops methods to model the second order behaviour of the GRF through considering both(i) oscillating precision structure, and (ii) non-stationarity. In the latter case two scenarios are considered:(i) spatially varying dependence due to some known effect, and (ii) spatially varying dependence due to some unknown effect. This is achieved through using both INLA and TMB (in situations where possible). Sections 7.2.1 and 7.2.2 detail the fitting of the above mention models in both `R-INLA` and TMB.

7.2.1 Oscillating precision matrix

This section considers an oscillating precision LGCP model. That is where the GRF has precision which varies in a repetitive manner. Specifically where the precision matrix $\mathbf{Q}$ is given by,

$$\mathbf{Q} = \kappa^4 \mathbf{C} + 2\kappa^2 \cos(\pi\,\theta)\,\mathbf{G}_1 + \mathbf{G}_2. \tag{7.3}$$

Here, $\theta = 0$ returns the regular Matérn case ($\cos(0) = 1$), and $0 < \theta < 1$ leads to an oscillating model, (for derivation see Lindgren et al. (2011)). Methods outlined below purely detail oscillating models with $0 < \theta < 1$. It should be noted that for $\theta = 1$, $\cos(\pi\,\theta) = -1$. In addition, for $n \in \mathbb{Z}$, $\cos(n\,\pi) = (-1)^n$; thus, the coefficient of $\mathbf{G}_1$ in the precision 7.3 alternates between $-2\,\kappa^2$ and $2\,\kappa^2$.

The left hand plot of Figure 7.2 shows the oscillating precision matrix, $\mathbf{Q}$, given by Equation (7.3). Specifically this is the precision of a GRF in a LGCP simulated in the unit square with parameterisation $\kappa = 12$ and $\theta = 0.9$.

7.2.2 Non-stationary spatial dependence

In this section the variance structure of the latent GRF, $\mathbf{z_s}$, is assumed to vary in space. In particular where a parameter, κ, of the precision matrix (Equation (7.2)), (i) depends on a set of known explanatory variables, and (ii) where the spatial dependence is estimated. The former is considered by Ingebrigtsen et al. (2014) where a regression model is put on a

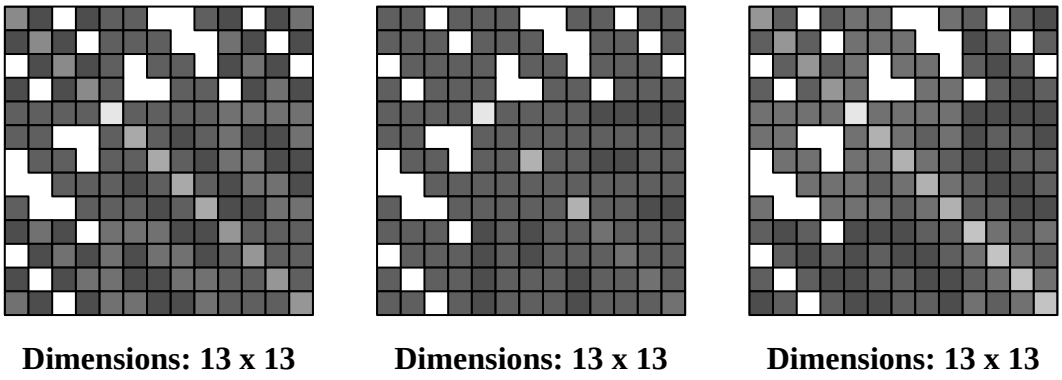


Figure 7.2: Plots illustrating three forms of the precision matrix $\mathbf{Q}$ for the GRF $\mathbf{z}$ simulated in the unit square. Here the shades of grey squares represent non-zero entries; where darker shades indicate larger values. Left hand plot shows an oscillating precision matrix given by Equation (7.3) with $\kappa = 12$ and $\theta = 0.9$. Centre plot shows a non-stationary precision matrix (Equation (7.4)) where $\text{diag}(\mathbf{D^1})_i = \exp(1+\sin(\pi\, x))$. Right hand plot shows a typical precision matrix as given by Equation (7.2) with $\kappa = 12$.

parameter of the precision matrix. In that case a regression model is put on the parameter τ (Equation (1.5)). In this section it is assumed that $\tau = 1$, and the spatial dependence is assumed to be reflected in the parameter κ (Equation (7.2)).

In order to account for this spatial dependence in the precision structure a generalisation of the precision matrix (Equation (7.2)) is required. This generalisation as given by Lindgren et al. (2011) is,

$$\mathbf{Q}_{ns} = \mathbf{D^1}\,\mathbf{C}\,\mathbf{D^1} + \mathbf{D^2}\,\mathbf{D^1}\,\mathbf{G}_1 + \mathbf{G}_1^T\,\mathbf{D^1}\,\mathbf{D^2} + \mathbf{G}_2. \tag{7.4}$$

As in Equation (7.2), $\mathbf{C}$, $\mathbf{G}_1$, and $\mathbf{G}_2$ are sparse matrices obtained from the finite basis representation Equation (1.7). The matrix $\mathbf{D^1}$ is diagonal where the diagonal elements are used to describe the non-stationary behaviour of κ, specifically,

$$\text{diag}(\mathbf{D^1})_i = \kappa_i = \exp(\theta_2 + \theta_3\, f(x_i, y_i)) + \epsilon_i. \tag{7.5}$$

The diagonal matrix $\mathbf{D^2}$ in order to be consistent with Lindgren et al. (2011) is in this section set to $\mathbf{D^2} = \exp(\,\mathrm{I}\,)$ (I is an identity matrix of dimension the same as $\mathbf{D^1}$). The centre plot of Figure 7.2 shows a non-stationary precision matrix, $\mathbf{Q}$. This being the precision of the GRF in a LGCP simulated in the unit square, given by Equation (7.4) where $\text{diag}(\mathbf{D^1})_i = \exp(1 + \sin(\pi\, x))$. This is similar to that considered by Krainski et al. (2016, Chapter 9).

The function $f(x_i, y_i)$ is a spatial function describing the spatial dependence of κ, which then has a direct effect on the marginal variance of the field σ^2. In particular, due to the relationship,

$$\sigma^2_{\mathbf{s}} = \frac{1}{4\,\pi\,\kappa^2_{\mathbf{s}}}.$$

One should note again the difference between oscillatory and non-stationary models, the

former do not affect the marginal variance, whereas the latter do through the relationship noted above.

The constraint of the methodology discussed above is the requirement of some known spatially varying variable with which κ_s is believed to vary. Yet, one may not know of any spatial function $f(x_i, y_i)$. In this instance the precision matrix (Equation (7.4)) remains the same, but may be thought of having

$$\text{diag}(\mathbf{D^1})_i = \kappa_i = \exp(u_i) + \epsilon_i, \tag{7.6}$$

for unknown $\mathbf{u} = (u_1, \cdots, u_n)$. This is akin to the regression given by Equation (7.5) with a spatially varying mean. This added level of randomness is clearly hugely computationally expensive, and not always possible due to the confounding of the variables in Equation (7.6).

The simulation and fitting of the models discussed in this and the previous section are implemented in the `R` package `lgcpSPDE` (detailed in Section 7.3). Both simulation and model fitting procedures rely on a SPDE model for the assumed GRF, which is constructed on the Delauney triangulation (Chapter 4). Fitting such LGCP models as outlined above is achieved through including additional arguments to the call `fit.lgcp()` (Section 4.3.1). Moreover implementation is facilitated using both `R-INLA` and TMB; Section 7.4 details a basic comparison of estimator performances for the above specified models.

7.3 Implementation in `lgcpSPDE`

Both the simulation of and fitting of the LGCPs mentioned above—exhibiting both types of discussed second order behaviour—are supported by `lgcpSPDE`. To simulate such models an additional argument—`non.stat`—must be supplied as a named list to the function `rlgcpspde()` (Section 4.3.2). Dependent on the type of model the user wishes to simulate the named elements differ. For example the following code (R Example 7.3.1) will result in LGCPs which have an oscillating and non-stationary second order behaviour respectively[1]. Here the objects `spatial.polygon` and `mesh.pars` are as those used in Section 4.3.2.

[1]Note by default `seed = 1` in the function `rlgcpspde()`

R Example 7.3.1.

```
OSC <- rlgcpspde(spatial.polygon = domain, mu = 1, mesh.pars = mesh.pars,
                 kappa = kappa, non.stat=list(oscillate = 0.6),
                 seed = seed)
#~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
ns.kappa <- rlgcpspde(spatial.polygon = domain, mesh.pars = mesh.pars,
                      mu = 1,
                      non.stat = list(fn = sin(pi*mesh$loc[,1]),
                                      theta = c(1,2,3))
```

In summary, the required elements which must be supplied when simulating either of the point pattern with second order behaviour which is either oscillatory, or non-stationary are:

- **Oscillatory:** `oscillate` a scalar specifying θ in Equation (7.3).
- **Non-stationary:** `fn` a function specifying the assumed spatially varying relationship between $\log(\kappa)$ and some fixed effect.
  `theta` a numeric vector of length three, `theta` $= (\theta^1, \theta^2, \theta^3)$ where θ_2 and θ_3 in Equation (7.5) are $\theta^2 - \theta^1$ and $\theta^3 - \theta^1$ respectively.

Estimation of the parameters associated with each model discussed herein requires only the supply of an additional argument to the function `fit.lgcp()` (Section 4.3). This argument—`ns`—must be a named list which must contain the (character) element `model` specifying the assumed model the user wishes to fit. This is restricted to one of the following,

- `lgcpTMB`: this will not fit a non-stationary model, but fits the equivalent to the LGCP model in Chapter 4 in TMB,
- `oscillateTMB`: an oscillating precision model using TMB,
- `nskappaTMB`: a non-stationary variance model using TMB (with user chosen spatial dependence),
- `nskappaINLA`: a non-stationary variance model using INLA (with user chosen spatial dependence),
- `nsUNkappaTMB`: a non-stationary variance model using TMB (with unknown second order dependence).

If one of the TMB models is chosen an additional element of the argument `ns` must be supplied; this being a list of parameter starting values which are supplied through the named list element as a named list, `parameters`. In each call fitting a TMB model the named

parameters `x` and `b0` must be supplied. The former a numeric vector of starting values for the GRF at the mesh nodes (i.e., length, `nv` in R Example 7.3.2). The latter a scalar giving the starting value of the intercept term β_0 (Section 4.3). Moreover, depending on the type of model to be fitted the named list `ns` must contain different named elements. R Example 7.3.2 illustrates how this argument would change to fit the above mentioned models respectively. Specifically, the additional (over `x` and `b0`) named parameter stating values supplied in each case are as follows,

- `lgcpTMB`: `log_kappa` starting value for the log of κ Equation (7.2) ,
- `oscillateTMB`: `log_kappa` starting value for the log of κ Equation (7.3), and `logit_theta` starting value for the logit of θ Equation (7.3),
- `nskappaTMB`: `theta2` and `theta3` scalars giving starting values for θ_2 and θ_3 in Equation (7.5),
- `nsUNkappaTMB`: `log_u`: a numeric vector of starting values for each u_i in Equation7.6 specified at the mesh nodes (i.e., length `nv` in R Example 7.3.2).

The final call to `fit.lgcp()` would in each case be `fit.lgcp(mesh = mesh, locs = locs, ns = ns)`.

It should be noted that the user must call the function `compile.lgcpSPDE()` prior to fitting any model using TMB. This as, TMB requires that some functionality is written as C++ type code. The package `lgcpSPDE` package contains the required templates, and thus any user of `lgcpSPDE` need not supply such. Yet the C++ templates must be compiled, which will be done automatically on the user's machine post running `compile.lgcpSPDE()`.

R Example 7.3.2.

```
ns <- list(parameters = list(beta0 = 2, log_kappa = 0,
                             x = rep(0,nv)),model="lgcpTMB")
#~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
ns.osc <- list(model = "oscillateTMB",
               parameters = list( log_kappa = 0.69,
                                  logit_theta = 0.4,beta0 = 1,
                                  x = rep(0,nv)))
#~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
ns.ns <- list(fn = sin(pi*mesh$loc[,1]),
              parameters = list(beta0 = 1, theta2 = 1, theta3 = 1,
                                x = rep(0,nv)), model = "nskappaTMB")
#~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
ns.inla <- list(fn = sin(pi*mesh$loc[,1]), model = "nskappaINLA")
```

Thus to fit an oscillating precision LGCP model to the point pattern object `OSC` in R Example 7.3.1 the user should call `fit.lgcp()` as below supplying a Delauney triangulation object (`mesh`) and the `ns` argument as `ns.osc` in R Example 7.3.2.

```
fit.oscillate <- fit.lgcp(mesh = mesh, locs = OSC, ns = ns.osc)
```

The object `fit.oscillate` is of class `sdreport` which is specific to fitting a TMB model. To obtain parameter estimates and standard errors the utility function `summary()` will work as usual, however to access specific estimates and additional argument to the `summary()` call should be given, as shown below.

```
summary(fit.oscillate,"fixed")
##              Estimate Std. Error
## log_kappa   0.1636778 0.14024569
## logit_theta 1.1051477 0.45389546
## beta0       0.9427801 0.09856267
summary(fit.oscillate,"report")
##        Estimate Std. Error
## theta 0.7512234 0.08482707
## kappa 1.1778347 0.16518625
```

7.4 Estimator performance

In order to ascertain estimation performance in the modelling of LGCP using the methodology described above, three small-scale simulation studies were carried out as follows:

1. Using TMB to fit a stationary LGCPs—`lgcpTMB`. That is, assuming a GRF with precision matrix given by Equation (7.2)

2. Using both TMB and INLA to fit non-stationary LGCPs—`nskappaTMB` & `nskappaINLA`. Specifically, assuming a GRF with precision matrix given by Equation (7.4) with $\text{diag}(\mathbf{D^1})_i$ being given by Equation (7.5).

3. Using TMB to fit non-stationary LGCPs—`nsUNkappaTMB`. Specifically, assuming a GRF with precision matrix given by Equation (7.4) with $\text{diag}(\mathbf{D^1})_i$ being given by Equation (7.6). It should be noted that the methodology associated with fitting such a model is very much in it's infancy. This chapter's main focus is to outline the concepts associated with non-stationary models.

Each study was carried out using functionality of the `R` package `lgcpSPDE` developed by the author of this thesis.

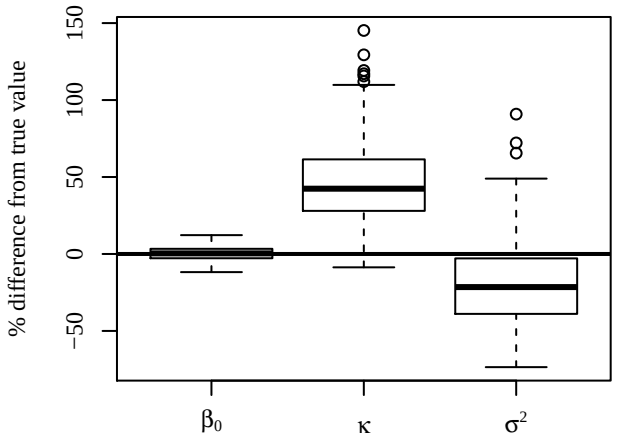


Figure 7.3: Boxplots of estimates showing the percentage difference from the true parameter values used to simulate the point pattern data ($\Lambda(\mathbf{s}) = \exp(\beta_0 + x(\mathbf{s};\theta))$, where $\beta_0 = 2$ and $\theta = (\kappa, \sigma^2) = (1, 0.05)$). There is comparatively higher variation, hence MSE (Table 7.1) for the (hyper) parameters κ and σ^2 of the latent field. This may be down to the resolution of the mesh, see Chapter 4.

	$Bias(\%)$	$RMSE(\%)$
β_0	0.22	0.18
κ	46.48	27.45
σ^2	-20.01	10.20

Table 7.1: Estimated biases, and root mean square errors from the simulation studies given as percentages of the true parameter values used to simulate the data.

The `lgcpTMB` study simulated a total of 500 point pattern data sets with parameterizarion $\Lambda(\mathbf{s}) = \exp(\beta_0 + x(\mathbf{s};\theta))$, where $\beta_0 = 2$ and $\theta = (\kappa, \sigma^2) = (1, 0.05)$. Estimator performance is shown in Figure 7.3 and Table 7.1. Here, the percentage differences from the true parameter values, and the percentage bias and standard deviations are shown respectively. Slight positive bias for the parameter β_0 (with little variance, hence, low MSE), and respective positive and negative bias for κ and σ^2 (both having larger variance, hence, large MSEs) is evident. This may be down to the resolution of the triangulation, mesh, see Chapter 4.

The non-stationary (models `nskappaTMB` and `nskappaINLA`) study simulated a total of 500 point pattern data sets with parameterisation $\Lambda(\mathbf{s}) = \exp(\beta_0 + x(\mathbf{s};\theta))$, where $\beta_0 = 1$, and $x(\mathbf{s};\theta)$ having precision given by Equation (7.5), with $\theta_2 = 2$, and $\theta_3 = 1$. Estimator performance is shown in Figure 7.4 and Table 7.2. The grey in Figure 7.4 indicate estimates obtained from the INLA-SPDE approach. There seem to be little difference between each estimator method. In both cases the estimators show little bias for the parameter β_0 (Table 7.1), however both the estimators show bias for the parameter θ_2 (positive using `R-INLA`, negative using TMB).

The `nsUNkappaTMB` study again simulated 500 point pattern datasets with parameterisation $\Lambda(\mathbf{s}) = \exp(\beta_0 + x(\mathbf{s};\theta))$, where $\beta_0 = 1$, and $x(\mathbf{s};\theta)$ having precision given by Equation (7.4) where $\text{diag}(\mathbf{D^1})$ is given by Equation (7.5). That is κ given in Equation (7.2) varies in space. The estimated bias—as a percentage—for β_0 was calculated as -5.38%, and the root mean squared error—again as a percentage—is calculated as 89.09%. Figure 7.5 illustrates, for one model fit—the estimated values of each κ_i at the mesh nodes—grey plotting char-

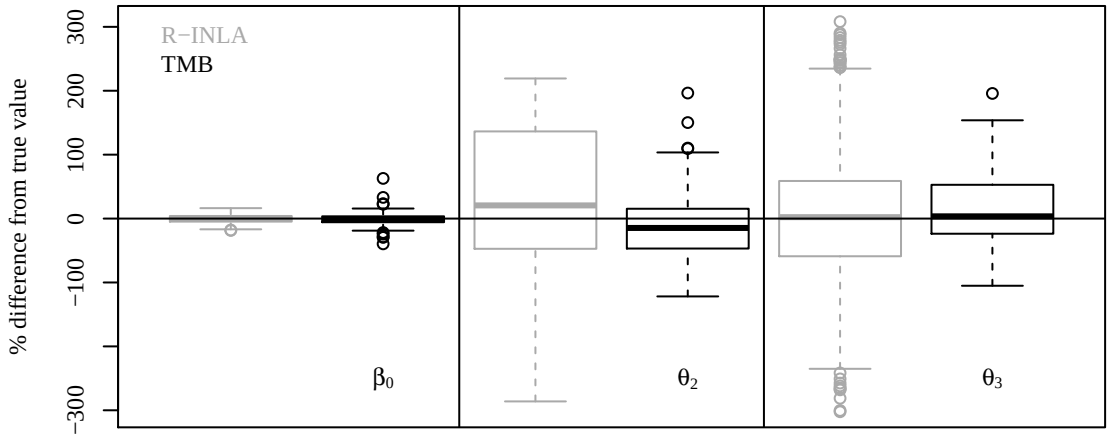


Figure 7.4: Boxplots of estimates showing the percentage difference from the true parameter values used to simulate the data. In particular, $\Lambda(\mathbf{s}) = \exp(\beta_0 + x(\mathbf{s};\theta))$, where $\beta_0 = 1$, and $x(\mathbf{s};\theta)$ having precision given by Equation (7.5), with $\theta_2 = 2$, and $\theta_3 = 1$. Grey relates to estimator `R-INLA`; black relates to estimator TMB.

Parameter	Estimator	$Bias(\%)$	$SD(\%)$	$RMSE(\%)$
β_0	INLA	-0.35	6.43	0.41
	TMB	-1.13	8.05	0.66
θ_2	INLA	29.81	110.20	130.33
	TMB	-14.44	43.97	21.42
θ_3	INLA	2.09	109.75	120.49
	TMB	12.26	53.39	30.01

Table 7.2: Estimator performance of `R-INLA` and TMB. Estimated biases, standard deviations, and root mean square errors from the simulation studies given as percentages of the true parameter values used to simulate the data.

acters. Larger plotting characters reflect larger values of κ_i. The black plotting characters in Figure 7.5 are the realisation of the simulated non-stationary LGCP to which the model was fitted. Recall that larger values of κ_i imply smaller marginal variace of the GRF, thus, lower spatial dependence. Assuming a spatially varying intercept model on κ results in a much larger computational burden. Moreover, the estimated values are confounded with the values of the estimated GRF. This as such methodology is akin to assuming two levels of randomness. Futher considerations of this methodolody is beyond the scope of this thesis, and left for later develompent by the authour. However, the reader should be able to envisage the flexibility that such a method offers. This as, the concept of assuming two levels of stochacisity (as above) would enable one to distuinguish between first and second order processes; thus, enabling one to infer the mechanisms which affect the realisation of the point process and the dependency structure of the point process.

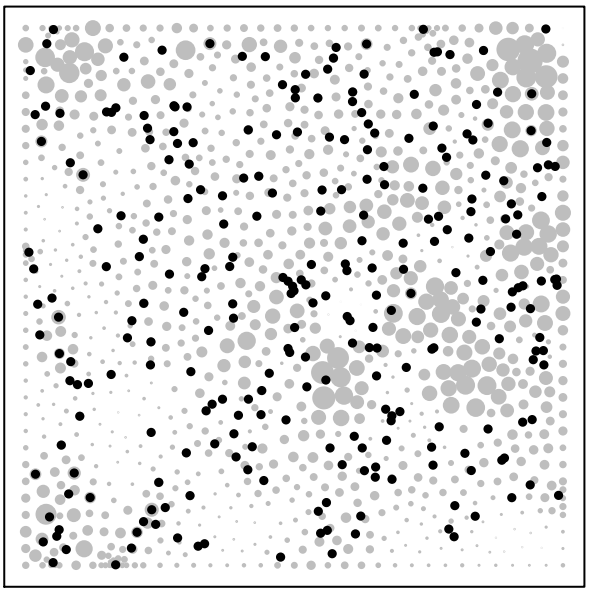


Figure 7.5: Illustration, for one fitted model, of the estimated values of each κ_i at the mesh nodes—grey plotting characters. Larger plotting characters reflect larger values of κ_i. Black plotting characters are the realisation of the simulated non-stationary LGCP.

7.5 Discussion

This chapter extends methodology presented in Chapters 4, 5, and 6; this through focusing on the second order behaviour of LGCPs. In addition, functionality of the `lgcpSPDE` package is extended to include utilising the estimator TMB in the fitting of those models discussed herein. The chief focus has been on the structure of the precision matrix of the GRF in each situation, and estimator performance. The estimation procedures—as in Chapters 4, 5, and 6—are centred around an SPDE model for the GRF.

This chapter aims at introducing the use of both estimators in fitting LGCP models, which are assumed to have second order behaviour as detailed herein. Specifically, this chapter highlights the change in parameterisation of the precision matrix when considering differences in the assumed GRF. That is, extensions of the typical precision matrix for the SPDE model for the GRF when considering oscillatory models (Equation (7.3)) and non-stationary behaviour (Equation (7.4)). Each extension considered assumes different second order behaviour of the GRF—the solution to the SPDE 1.5. Such behaviour has typically been explored in reference to environmental data. However it may be the case in many real world phenomena that the true dependence structure is spatially varying. For brevity, from this point on, spatially varying dependence is used to refer to both non-stationary and oscillatory behaviour (noted as being distinct above). In this case assuming a non spatially varying dependence structure in one's model is inadequate; in this case inference may be affected. Despite this, allowing more flexibility in the GRF may in some instances

be unnecessary, and lead to overfitting. This concept is addressed by Fuglstad et al. (2015a) in the context of geostatistical data. In particular, they note that determining the form of the spatially varying dependence is paramount, as allowing over flexibility in the GRF may lead to spurious inference. Thus, in order to reflect second order behaviour adequately the required flexibility of the model should be available where appropriate.

This chapter—as has this thesis—focused on the structure of the GRF in fitting LGCPs. Throughout, the SPDE model—Equation (1.5)—for the GRF has remained unchanged. It is only under certain assumptions, outlined in Section 1.3.2, that the solution to Equation (1.5) is a GRF with precision matrix given by Equation (7.2). Yet, there is scope in extending Equation (1.5) to describe a more complex process (e.g., non-separable space-time). Comprehensively exploring these extensions are outwith the aims of this thesis, but a few possible extensions are detailed in the following Discussion (Chapter 8).

Part VI

Discussion

Chapter 8

Discussion

Each chapter of this thesis details methods to fit point process models, primarily LGCPs. Parameter estimation is accomplished through either an approximate likelihood approach (Chapters 2 and 3), a Bayesian approach (Chapters 4, 5 and 6), or an AD approach (Chapter 7). Recent years have seen a huge development in computationally efficient methods designed to fit spatial models (e.g., INLA and TMB), and such methodology enables the modelling of more complex spatio-temporal dependency structures. These developments are particularly useful when considering mark-point dependence in a marked LGCP, as in Chapter 5, or multi-species dependencies, as in Chapter 6. In addition to modelling the first order structure of the assumed process (e.g., intensity measure of a LGCP), Chapter 7 demonstrates how further behaviour of the assumed stochastic structure can be accounted for.

There is great scope for the further developments of the point process modelling methods discussed herein, particularly in the context of LGCPs, where the shared latent GRFs account for dependencies in multivariate data. Part IV demonstrates the flexibility of this methodology in its application within the fields of ecology, and geology. Moreover, the joint models implemented in the `lgcpSPDE` package are currently being used modelling disease incidence in Namibia where there is covariate misalignment (Dismas Ntirampeba, personal communication, September 2016). This work aims to account for local socio-economic, and environmental factors in modelling the spatial distribution of incidence of HIV. This, when the data for disease incidence, socio-economic, and environmental covariates are collected at different spatial scales.

The role of the GRF, modelled by a SPDE, is an integral part of Chapters 4 onward. In each chapter the SPDE model is assumed to be of the form given in Equation (1.5). This is a requirement for the solution of Equation (1.5) to be a GRF with Matérn covariance. However, other structures of latent processes may be a better reflection of the mechanisms operating—specifically, co-dependence structures operating between space and time.

The following sections in this chapter discuss methods that consider different forms of the SPDE given by Equation (1.5). Section 8.1 considers a spatio-temporal processes. Section

8.2 considers a bivariate system of SPDEs with a spatial lag. Both of the above mentioned extensions are in addition to those developed in the chapters above, and individually, or jointly provide an extremely flexible class of point process model. This is achieved through the conjectures made regarding space time interactions (Section 8.1), or bivariate GRFs (Section 8.2). The former case is an extension to the spatio-temporal models discussed in Chapters 5, and 6, which purely considers spatial processes progressing through time, without any space time interaction. The latter case discusses methodology akin to that developed in Chapters 5, and 6 regarding multivariate GRFs.

8.1 Spatio-temporal processes

Typically, spatio-temporal data are dealt within a hierarchical statistical framework (Banerjee et al., 2008), whereby a latent process is assumed to drive the observations (as in all Chapters of this thesis). Due to favourable analytic properties, this process is usually considered to be a continuously indexed GRF which is defined through a mean function $\mu(.)$, and some covariance function $C(.,.)$. The following section discusses a situation where there is some space-time interaction in the GRF, initially through considering a non-separable covariance, and then considering a spatial-temporal SPDE. These sections are purely speculative in nature, aimed at simply proposing extensions of the methodology considered in this thesis.

8.1.1 Nonseparable spatio-temporal covariance

One of the main complexities associated with spatio-temporal processes is the specification of an appropriate covariance function that adequately captures the true spatio-temporal dependence inherent in the data (Stein, 2005; Cressie and Huang, 1999; Gneiting, 2002; Fuentes et al., 2008).

Consider data that have both a spatial and temporal index, $Y(\mathbf{x}_{1i}, t_i), ..., Y(\mathbf{x}_{n_i i}, t_i) : i = 1, ..., m$, where $\{\mathbf{x}_{1i}, ..., \mathbf{x}_{n_i i}\} : \mathbf{x} \in \mathbb{R}^{\mathrm{d}}$ are the n_i data locations at each time point i, with $\{t_1 < t_2 < ... < t_m\} : t \in \mathbb{R}$ being the times of observation. A typical hierarchical model formulation for such spatio-temporal data is given by,

$$\begin{aligned} Y(\mathbf{x}, t) &= \mu(\mathbf{x}, t) + \epsilon(\mathbf{x}, t) \\ \mu(\mathbf{x}, t) &= Z(\mathbf{x}, t) + \varepsilon(\mathbf{x}, t), \end{aligned} \tag{8.1}$$

where $\varepsilon(\mathbf{x}, t)$ is a Gaussian white noise process and $Z(\mathbf{x}, t)$ is a mean-zero spatio-temporal process.

Using such a model formulation given by Equation (8.1) the interest is in inferring both the spatial and temporal dependence in the spatio-temporal process. One important and intuitive point of note is that although mathematically $\mathbb{R}^d \times \mathbb{R} = \mathbb{R}^{d+1}$ this is clearly not the case when considering d dimensional space and (1D) time. Simply put, an observation at time

t, at 2D location (x, y) is not an observation in some 3D space. This, due to the disparity in dimensional measurements. It is assumed that one can move freely in any spatial direction, whereas one can only (subjectively) travel forward in time. The three main approaches which mathematically reflect the difference between space and time are (i) including some scaling parameter that scales the temporal units allowing one to properly account for the difference in measurement units; (ii) assume space time additivity, that is, $\text{cov}(Y(\mathbf{x}, t)) = \mathcal{C}^s(\mathbf{x}, \mathbf{t}) + \mathcal{C}^t(t, r)$; or (iii) assume space-time multiplicity, that is, $\text{cov}(Y(\mathbf{x}, t)) = \mathcal{C}^s(\mathbf{x}, \mathbf{t})\,\mathcal{C}^t(t, r)$. Where $\mathcal{C}^s$ and $\mathcal{C}^t$ are spatial and temporal covariance functions respectively. In the (separable) approaches (ii) and (iii), no account is given for space-time interaction, which, in practice, is often ignored due to the added complexity.

Typically, assuming a separable covariance is much simpler, due to their explicit interpretation and relative computational inexpensiveness, and as such their use is much more prevalent in the context of spatio-temporal statistics (Cressie and Wikle, 2011). However, separability may not always be an appropriate assumption for the covariance of some spatio-temporal processes. Many nonseparable space-time covariance functions have been previously considered to account for the space-time interaction. In such cases a parameter is estimated to measure the strength of the space-time interaction based on the spectral representation of the process (Cressie and Huang, 1999; Stein, 2005; Fuentes et al., 2008). Gneiting (2002) constructed a class of non-separable space-time covariance functions directly from the spatial, $\mathbf{x}$, and temporal, t, domains which are of the form $C(\mathbf{x}; t) = \frac{\sigma^2}{\psi(|t|^2)}\phi(\frac{\|\mathbf{x}\|^2}{\phi(|t|^2)})$, where $(\mathbf{x}; t) \in \mathbb{R}^d \times \mathbb{R}$. Here $\phi(\cdot)$ is a monotone function and $\psi(\cdot)$ is positive and has a completely monotone derivative, with argument ≥ 0 .

Following the procedure outlined below, put forward by Särkkä and Hartikainen (2012), a SPDE model can directly be derived corresponding to a covariance function of choice. In summary, Särkkä and Hartikainen (2012) propose the following steps:(i) compute the spectral density via a Fourier transform of the covariance function, (ii) form the rational approximation in $i\mathbf{w}$, and (iii) convert this into a state-space equation to form the required SPDE.

Therefore, following this procedure, the remainder of this section considers a nonseparable Matérn type covariance, and the corresponding spectral density (which is simply the Fourier transform of the aforementioned covariance). This spectral density represents the stochastic process with the pre defined covariance.

- A nonseparable Matérn type covariance:

  Consider a 3D Matérn type covariance function which has an extra parameter to account for the difference in temporal and spatial unit measurements given by

  $$C(\mathbf{x}, t) = C(\mathbf{x}^*) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa\|(\mathbf{x}, \rho t)\|)^\nu K_\nu(\kappa\|(\mathbf{x}, \rho t)\|), \tag{8.2}$$

  where κ is a scaling parameter, ν is a smoothness parameter, K_ν is the modified Bessel function, and ρ is some temporal scaling parameter which allows us to take into account

the change of units between the spatial and temporal domains. Equation (8.2) is therefore a 3D Matérn type spatio-temporal covariance, with $\|(\mathbf{x}, \rho t)\|$ being a Euclidean type distance metric, leading to an anisotropic field.

- Spectral density:

  Spectral density is often used to represent the assumed stochastic process. It shows the strength of the variations as a function of frequency. In the case discussed here, the stochastic process represented is spatio-temporal with a nonseparable Matérn type covariance. Here the term frequency is some transformation of space or time. Thus, looking at variations in the frequency domain is just another way to look at variations in either space or time of the data—akin to a latent field. Therefore, modelling the nonseparable spatio-temporal field represented by the above spectral density would enable the space-time interaction inherent in many spatio-temporal data to be modelled.

  The spectral density $S(\mathbf{w_x}, w_t)$ of $Z(\mathbf{x}, t)$ changes with space and time to explain how the spatial temporal dependency varies on the domain of interest. A special case of the stationary spectral density proposed by Fuentes et al. (2008) is

$$S(\mathbf{w_x}, w_t) = \frac{\gamma}{(\alpha^2\beta^2 + \alpha^2\|\mathbf{w_x}\|^2 + \beta^2\|w_t\|^2 + \varepsilon\|\mathbf{w_x}\|^2\|w_t\|^2)^\nu}, \tag{8.3}$$

  where α is the rate of decay of the spatial correlation, β is the rate of decay of the temporal correlation, $\nu > (d+1)/2$, and ε controls the separability of the process—the interaction between space and time in the model. For $0 \leq \epsilon < 1$ the corresponding process is nonseparable, and separable for $\epsilon = 1$.

  The stationary spectral density, $S(\mathbf{w_x}, w_t)$, by setting $\varepsilon = 0$ in Equation (8.3), is an extension to the traditional Matérn spectral density. Thus, the spectral density of Equation (8.2) is given by

$$S(\mathbf{w_x}, w_t) = \frac{\gamma}{(\alpha^2\beta^2 + \alpha^2\|\mathbf{w_x}\|^2 + \beta^2\|w_t\|^2)^\nu}.$$

Figure 8.1 illustrates the spectral density given by Equation (8.3) where $\gamma = \alpha = \beta = 1$ and $\nu = \frac{3}{2}$ for varying values of ε. As $\varepsilon \to 1$ the ridges of the spatio-temporal covariance become increasingly pronounced. Thus, assuming a stochastic process represented by the spectral density detailed above, would enable estimation of the strength of the space time interaction that may be present in the data. This flexibility would be exceptionally useful when attempting to infer phenomena which do not occur independently in space and time.

8.1.2 A space-time random field

A further method to consider a nonseparable space-time process would be to consider a spatio-temporal SPDE of the form,

$$(\alpha^2 - \frac{\alpha^2}{\beta^2}\Delta)f(\mathbf{x}, t) + \frac{d}{dt}f(\mathbf{x}, t) = w(\mathbf{x}, t). \tag{8.4}$$

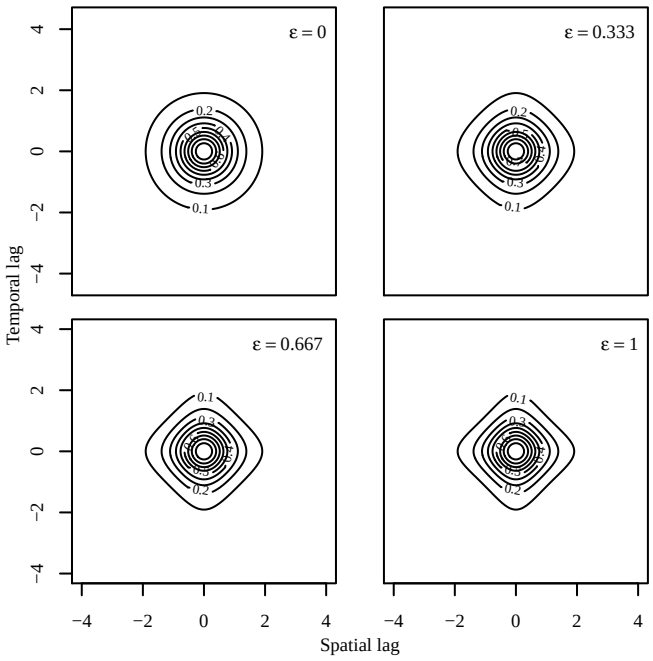


Figure 8.1: Spatio-temporal spectral densities as per Equation (8.3) with $\gamma = \alpha = \beta = 1$ and $\nu = 1.5$ for different ε. For $\varepsilon \neq 0 \to 1$ the ridges of the spatio-temporal covariance become increasingly pronounced. This as for $0 \leq \varepsilon < 1$ the spectral density (Equation (8.3)) is nonseparable and separable for $\varepsilon = 1$.

Here $w(\mathbf{x}, t)$ is spatio-temporal white noise, Δ is the Laplacian, and (as in the section above) α^2/β^2 is a ratio of the spatio and temporal scaling parameters.

Considering the spatio-temporal SPDE given by Equation (8.4), this section goes on in a similar vein to Lindgren et al. (2011) to derive the precision matrix of the spatio-temporal process (i.e., a spatio-temporal GRF).

Considering the SPDE given by Equation (8.4), a two-dimensional spatial domain (d= 2) is assumed throughout this section. Following the procedures, and using the definitions given by Lindgren et al. (2011) the elements of the precision matrix of the Markov representation of the nonseparable space-time GRF, $f(\mathbf{x}, t) = f(\mathbf{x}^*)$, can be derived. From the basis representation given in Equation (1.7) (Section 1.3.2) the weight vector $\mathbf{w}$ is $\mathbf{N}(\mathbf{0}, \mathbf{Q^{-1}})$, where now $f(\mathbf{u}) = f(\mathbf{x}, t) = f(\mathbf{x}^*)$.

Considering a SPDE of the form

$$\mathbb{L}\ f(\mathbf{x}, t) = w(\mathbf{x}, t), \quad \text{when} \quad \mathbb{L} = \alpha^2 - \tfrac{\alpha^2}{\beta^2}\Delta + \tfrac{d}{dt},$$

defining the inner product,$\langle \cdot, \cdot \rangle$, as,

$$\langle g, h \rangle = \int_{\mathbb{R}^1} \int_{\mathbb{R}^2} g(\mathbf{x}, t)^T h(\mathbf{x}, t) d\mathbf{x} dt,$$

and through defining the following matrices:

$$\begin{aligned}
\mathrm{C}_{ij} &= \langle \phi_i, \phi_j \rangle \\
\mathrm{G}_{ij} &= \langle \nabla\phi_i, \nabla\phi_j \rangle \\
\mathrm{K}_{ij} &= (\alpha^2 + 1)\mathrm{C}_{ij} + \frac{\alpha^2}{\beta^2}\mathrm{G}_{ij},
\end{aligned}$$

the following conjecture is made.

Conjecture:

Let $\mathbf{w_g}, \mathbf{w_h}$ be the Hilbert space coordinates of the two test functions g_n and h_n, then,

$$\begin{aligned}
\langle g_n, \mathbb{L}f_n \rangle_\Omega &= \Sigma_{ij} w_{g,i} \langle \phi_i, \mathbb{L}\phi_j \rangle_\Omega w_j \\
&= \Sigma_{ij} w_{g,i} (\alpha^2 + 1) C_{ij} + \frac{\alpha^2}{\beta^2} G_{ij} w_j \\
&= \mathbf{w_g}^T \mathbf{K} \mathbf{w},
\end{aligned}$$

due to Green's identity. Moreover,

$$\begin{aligned}
\mathrm{cov}(\langle g_n, \mathbb{L}f_n \rangle_\Omega, \langle h_n, \mathbb{L}f_n \rangle_\Omega) &= \mathbf{w_g}^T \mathbf{K} \mathrm{cov}(\mathbf{w}, \mathbf{w}) \mathbf{K}^T \mathbf{w_h} \\
&= \mathrm{cov}(\langle g_n, \mathbf{W} \rangle_\Omega, \langle h_n, \mathbf{W} \rangle_\Omega) \\
&= \langle g_n, h_n \rangle_\Omega \\
&= \Sigma_{ij} w_{g,i} \langle \phi_i, \phi_j \rangle_\Omega w_{h,j} \\
&= \Sigma_{ij} w_{g,i} C_{ij} w_{h,j} \\
&= \mathbf{w_g}^T \mathbf{C} \mathbf{w_h}
\end{aligned}$$

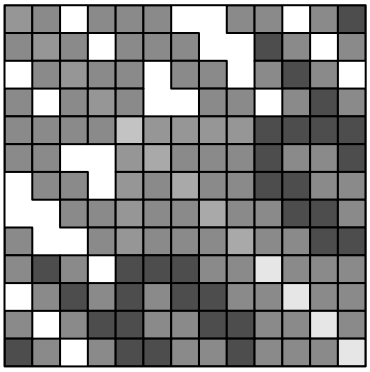
for every pair of test functions (g_n, h_n). In this situation the following precision matrix for the spatial-temporal solution to the SPDE given by Equation (8.4) is obtained

$$\begin{aligned}
\mathrm{cov}(\mathbf{w}, \mathbf{w}) = \mathbf{K}^{-1}\mathbf{C}\mathbf{K}^{-T}&, \text{ i.e., when} \\
\mathbf{Q} = \mathbf{K}^T \mathbf{C}^{-1} \mathbf{K}&
\end{aligned}$$

In keeping with the form of the precision matricies discussed in Chapter 7, this precision is written as

$$\begin{aligned}
\mathbf{Q} &= \mathbf{K}^T \mathbf{C}^{-1} \mathbf{K} \\
&= \left((\alpha^2 + 1)\mathbf{C} + \tfrac{\alpha^2}{\beta^2}\mathbf{G}\right)^T \mathbf{C}^{-1} \left((\alpha^2 + 1)\mathbf{C} + \tfrac{\alpha^2}{\beta^2}\mathbf{G}\right) \\
&= \left((\alpha^2 + 1)\mathbf{C^T} + \tfrac{\alpha^2}{\beta^2}\mathbf{G^T}\right) \mathbf{C}^{-1} \left((\alpha^2 + 1)\mathbf{C} + \tfrac{\alpha^2}{\beta^2}\mathbf{G}\right) \\
&= \left((\alpha^2 + 1)\mathbf{C^T}\mathbf{C}^{-1} + \tfrac{\alpha^2}{\beta^2}\mathbf{G^T}\mathbf{C}^{-1}\right) \left((\alpha^2 + 1)\mathbf{C} + \tfrac{\alpha^2}{\beta^2}\mathbf{G}\right) \\
&= (\alpha^2 + 1)^2\mathbf{C} + 2\,\tfrac{\alpha^2}{\beta^2}\mathbf{G} + \tfrac{\alpha^4}{\beta^4}\mathbf{G}\mathbf{C^{-1}}\mathbf{G}
\end{aligned} \tag{8.5}$$

as both $\mathbf{G}$ and $\mathbf{C}$ are symmetric. Figure 8.2 illustrates the possible form of this precision matrix, where $\alpha^2 = 4$ and $\beta^2 = 2$. As per the methods discussed in Chapter 7 defining



Dimensions: 13 x 13

Figure 8.2: An illustration of the precision matrix given by Equation (8.5) with $\alpha^2 = 4$ and $\beta^2 = 2$. Here the shades of grey squares represent non-zero entries; where darker shades indicate larger values.

the precision matrix of a GRF is an integral part in modelling the structure of the latent mechanisms in a point process. Moreover, in order that the computational burden typically associated with spatio-temporal processes be circumvented, this precision should be sparse. Thus, methodology which adequately captures the interaction between space and time in a computationally feasible way is highly desirable.

8.2 A spatial lag in a system of Gaussian random fields

In many real world situations there exist examples of more than one point process operating in the same area. Examples of this may be two competing species in the same habitat, or the locations of crimes and police presence in the same city. In these situations the latent spatial fields assumed for each process are likely to be linked. This section considers the bivariate system of SPDEs discussed by Hu et al. (2013). This system of SPDEs may be used to construct a bivariate random field which can be thought of as the latent field in a bivariate point process model.

In addition, an extension to the system introduced by Hu et al. (2013) is discussed. Specifically, through the inclusion of a spatially lagged link describing the dependence between two latent fields. Classically, one may consider modelling each process independently; however, this would be inadequate in many situations where there is some dependence between the fields.

Utilising this approach in constructing bivariate GRFs from an application point of view aids in the avoidance of the computational burden associated with defining multivariate dependence—this is due to maintaining the sparseness properties of the precision matrix. Moreover, the concept of some spatial lag, δ, leads to a nonstationary bivariate system of GRFs (Chapter 7 considered only a univariate system).

Letting L_{ij} for $i, j = 1, 2$ be differential operators and let $X(s) = (x_1(s), x_2(s))^T$ be the solution to a system of bivariate SPDEs given by

$$\begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \tag{8.6}$$

where each f_i is a white noise like process. Hu et al. (2013) consider in detail the following system which they termed the triangular version of the bivariate system (8.6):

$$\begin{aligned} b_{11}(\kappa_{11}^2 - \Delta)^{\frac{\alpha_{11}}{2}} x_1(s) &= f_1(s) \\ b_{22}(\kappa_{22}^2 - \Delta)^{\frac{\alpha_{22}}{2}} x_2(s) + b_{21}(\kappa_{21}^2 - \Delta)^{\frac{\alpha_{21}}{2}} x_1(s) &= f_2(s) \end{aligned} \tag{8.7}$$

The solution to the system given in Equation (8.7) is a bivariate random field denoted by $X(s) = (x_1(s), x_2(s))^T$, the parameters $\{b_{ij};\, i, j = 1, 2,\, i > j\}$ control the correlation between the two fields (i.e., the parameter b_{21} in Equation (8.7)). Hu et al. (2013) also show that each b_{ij} relate to the variance of the GRFs, each κ_{ij} relates to the range of the two fields, and each α_{ij} relates to the smoothness of the two fields. The remainder of this section discusses the case where the b_{21} parameter in the system (8.7) is spatially dependent, more specifically the two fields are related through some spatial lag δ (i.e., the cross-covariance is spatially dependent). More specifically, the aim is to explore what mechanisms operating in physical world phenomena could be captured if b_{21} were to be related to a spatial lag operator, δ. This would be pertinent, if, for example, one GRF were related to another through some process based on spatial lag (i.e., consider one such process mimicking another operating at some lag δ).

In this case the bivariate random field would be represented by the following system of SPDEs:

$$\begin{aligned} (\kappa_{n1}^2 - \Delta)^{\frac{\alpha_{n1}}{2}} f_1(s) &= W_1(s) \\ (\kappa_{n2}^2 - \Delta)^{\frac{\alpha_{n2}}{2}} f_2(s) &= W_2(s) \\ b_{11}(\kappa_{11}^2 - \Delta)^{\frac{\alpha_{11}}{2}} x_1(s) &= f_1(s) \\ b_{22}(\kappa_{22}^2 - \Delta)^{\frac{\alpha_{22}}{2}} x_2(s) + b_{21}(s)(\kappa_{21}^2 - \Delta)^{\frac{\alpha_{21}}{2}} x_1(s) &= f_2(s), \end{aligned} \tag{8.8}$$

where Hu et al. (2013) showed that for $\kappa_{n1} = \kappa_{11}$ and $\kappa_{n2} = \kappa_{22}$, $x_1(s)$ was a Matérn random field, and from Bolin and Lindgren (2011), $x_2(s)$ is close to a Matérn field. In Equation (8.8) $b_{21}(s)$ captures the dependence between the two fields and is related to a spatial lag δ.

As described in detail by Hu et al. (2013), the bivariate system of GRFs (8.7) where $\mathbf{X}(s) = (x_1(s), x_2(s))^T$ has the $2N$ dimensional multivariate Gaussian distribution given by

$$\pi(\mathbf{X}|\theta) = (\frac{1}{2\pi})^{2N} \|\mathbf{Q}(\theta)\|^{\frac{1}{2}} \exp[-\frac{1}{2}\mathbf{X}^T \mathbf{Q}(\boldsymbol{\theta})\mathbf{X}],$$

where $\mathbf{Q}(\theta)$ is the precision matrix of the bivariate GMRF with parameters $\boldsymbol{\theta}$, given by

$$\mathbf{Q} = \mathbf{K}\mathbf{D}^{-1}\mathbf{Q}_\mathbf{f}\mathbf{D}^{-1}\mathbf{K}, \tag{8.9}$$

Where $\mathbf{D}, \mathbf{Q_f}$ and $\mathbf{K}$ are defined by

$$\mathbf{D} = \begin{pmatrix} \tilde{\mathbf{C}} & \\ & \tilde{\mathbf{C}} \end{pmatrix}, \mathbf{K} = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}, \mathbf{Q_f} = \begin{pmatrix} Q_{f1} & \\ & Q_{f2} \end{pmatrix},$$

with $K_{ij} = b_{ij}(\kappa_{ij}^2 \tilde{C} + G)$, $\tilde{C_{ii}} = \langle \psi_i, 1 \rangle$, $G_{mn} = \langle \nabla\psi_m, \nabla\psi_n \rangle$, for $i, j = 1, 2$ and $m, n = 1, 2, ..., N$. The precision matricies for the noise processes $f_i(s)$ are $Q_{\kappa_{ni}^2} = K_{\kappa_{ni}^2}^T \tilde{C^{-1}} K_{\kappa_{ni}^2}$ for $\alpha_{ij} = 2$, where each ψ_i is a basis function in Equation (1.7).

Note that the precision matrix given by Equation (8.9) for the bivariate GMRF is linked to the $b'_{ij}s$ through the $\mathbf{K}$ matrix. Now, recall that Equation (8.7) only considers the triangular version of SPDEs, that is, $i, j = 1, 2$, for $i > j$. Hence $\mathbf{K}$ becomes,

$$\mathbf{K} = \begin{pmatrix} K_{11} & \\ K_{21} & K_{22} \end{pmatrix}.$$

Here only K_{21} is spatially dependent through $b_{21}(s)$ being related to some spatial lag δ, (i.e., $K_{21} = b_{21}(s)(\kappa_{21}^2 \tilde{C} + G)$. The consequence of this is a non-stationary bivariate GRF, as each b_{ij} is related to the marginal variance of the GRF. This reflects that the marginal variance of each GRF solution to Equation (8.8) not only depends on spatial location but the inter-dependence between the GRFs. Due to time constraints, no further development has been made by the author of this thesis into exploring this concept. However, it is clear that the development of non-stationary bivariate—or indeed multivariate—GRFs would establish a more flexible class of point process model enabling a computationally efficient technique by which to capture the complex dependencies inherent in co-dependent processes.

8.3 Concluding remarks

Two variants of the SPDE given by Equation (1.5) have been discussed over the previous two sections, and in particular, their applicability in modelling point processes. Such systems as described by both Section 8.1 and 8.2 have rarely been implemented in relation to point processes; tending to remain explored in the perhaps less applied literature alone (Illian et al., 2012a; Bolin and Lindgren, 2011; Hu et al., 2013; Brix and Diggle, 2001). However, the flexibility offered by each considered extension is hugely important to the fitting of models that are able to adequately reflect the structure of real world data. In particular, methods which are able to model spatio-temporal interaction in a computationally feasible, and meaningful way, would circumvent the misguided oversimplification of spatio-temporal processes. Moreover, being able to estimate dependencies, either between spatial processes, or interactions between space and time, would lead to a generalised modelling framework of which the latent Gaussian models discussed in this thesis are special cases.

In addition to the extensions mentioned above, with regards to the point process methodologies detailed in this thesis, there are a few more points of interest to note. Under current

development with regards to the INLA framework are methods assessing goodness of fit. Goodness of fit in this framework is currently based on the DIC criterion (Spiegelhalter et al., 2002), or predictive cross-validation measures (Held et al., 2010); yet, this does require further development in reference to point process models. While prior sensitivity analysis is not explicitly discussed in this thesis (other than by passing in Chapter 6), choice of priors for INLA-based models is still an area which requires further scrutiny (Sørbye and Rue, 2014; Fuglstad et al., 2015b) over and above the penalised complexity priors detailed in Simpson et al. (2014) and Martins et al. (2014).

This thesis has described methods by which to fit point process models (e.g., approximate maximum likelihood, Bayesian approach based on Laplace approximations, and an approach based on AD and Laplace approximation). In addition, the flexibility of the SPDE model negating any loss of spatial information is discussed. However, the main focus of this thesis was the methods developed within that enable the modelling of complex dependencies in both marked point pattern, and multivariate data.

Point process models are becoming increasingly commonplace (Møller and Waagepetersen, 2007); (Zammit-Mangion et al., 2012);(Python et al., 2016); Stevenson et al. (*in submission*); Jones-Todd et al. (*in submission*). Yet, few approaches consider utilising the flexibility that GRFs offer, in particular, when there are possible unknown dependencies or abstract structures inherent in the data. Modelling some unobserved process, in many cases, is apt in order to capture the driving mechanisms. Moreover, with the increasing advancement of computational capabilities the progression of point process models is cementing their versatility, and practicability within various scientific fields.

Bibliography

Aitchison, J. (1955). On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the American Statistical Association*, 50(271):901–908.

Albertsen, C. M., Whoriskey, K., Yurkowski, D., Nielsen, A., and Mills, J. (2015). Fast fitting of non-Gaussian state-space models to animal movement data via Template Model Builder. *Ecology*, 96(10):2598–2604.

Assunção, R. M., Reis, I. A., and Oliveira, C. D. L. (2001). Diffusion and prediction of leishmaniasis in a large metropolitan area in Brazil with a Bayesian space–time model. *Statistics in Medicine*, 20(15):2319–2335.

Baddeley, A. and Turner, R. (2005). `spatstat`: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848.

Barthelmé, S., Trukenbrod, H., Engbert, R., and Wichmann, F. (2013). Modeling fixation locations using spatial point processes. *Journal of Vision*, 13(12):1–34.

Benes, V., Bodlák, K., Møller, J., and Waagepetersen, R. P. (2002). Bayesian analysis of log Gaussian Cox processes for disease mapping. Technical report, Aalborg Universitetsforlag.

Bibby, C., Burgess, N., Hill, D., and Mustoe, S. (2000). *Bird Census Techniques*. Academic Press, London, 2nd edition.

Blangiardo, M. and Cameletti, M. (2015). *Spatial and spatio-temporal Bayesian Models with R-INLA*. John Wiley & Sons, Chichester.

Bolin, D. and Lindgren, F. (2011). Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *The Annals of Applied Statistics*, 5(1):523–550.

Boulangeat, I., Gravel, D., and Thuiller, W. (2012). Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecology Letters*, 15(6):584–593.

Brewer, R. and Sleeman, J. (1960). Soil structure and fabric. *Journal of Soil Science*, 11(1):172–185.

Brix, A. and Diggle, P. J. (2001). Spatiotemporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4):823–841.

Caie, P. D. and Harrison, D. J. (2016). *Next-Generation Pathology*. Springer, New York.

Caie, P. D., Turnbull, A. K., Farrington, S. M., Oniscu, A., and Harrison, D. J. (2014). Quantification of tumour budding, lymphatic vessel density and invasion through image analysis in colorectal cancer. *Journal of Translational Medicine*, 12(156):1–12.

Cannon, A. H., Cowen, L. J., and Priebe, C. E. (1998). Approximate distance classification. In *Computing Science and Statistics: Proceedings of the 30th Symposium on the Interface*, pages 544–549. Citeseer.

Clark, J. S. (2003). Uncertainty and variability in demography and population growth: a hierarchical approach. *Ecology*, 84(6):1370–1381.

Cox, D. R. and Isham, V. (1980). *Point processes*. Chapman and Hall, London.

Cressie, N. (1993). *Statistics for Spatial Data*. Wiley-Interscience, New York.

Cressie, N. and Huang, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94(448):1330–1339.

Cressie, N. and Wikle, C. K. (2011). *Statistics for spatio-temporal data*. John Wiley & Sons, Hoboken.

DeVinney, J. and Priebe, C. E. (2006). A new family of proximity graphs: class cover catch digraphs. *Discrete Applied Mathematics*, 154(14):1975–1982.

Diggle, P. and Ribeiro, P. J. (2007). *Model-based geostatistics*. Springer, New York.

Diggle, P. J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns*. CRC Press, Boca Raton.

Diggle, P. J. and Milne, R. K. (1983). Bivariate Cox processes: some models for bivariate spatial point patterns. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 45(1):11–21.

Diggle, P. J., Tawn, J., and Moyeed, R. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350.

Dobrỳ, M. (2011). The abundance of the little owl (*Athene noctua*) in Podunajská rovina lowland in 2009 and 2010. *Slovak Raptor Journal*, 5(1):121–126.

Dray, S., Pélissier, R., Couteron, P., Fortin, M.-J., Legendre, P., Peres-Neto, P., Bellier, E., Bivand, R., Blanchet, F. G., De Cáceres, M., Dunfout, A., Heegaard, E., Jombart, T., Munoz, F., Oksanen, J., Thioulouse, J., and Wager, H. (2012). Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs*, 82(3):257–275.

Fewster, R. M., Stevenson, B. C., and Borchers, D. L. (2016). Trace-contrast models for capture–recapture without capture histories. *Statistical Science*, 31(2):245–258.

Foster, S. D. and Bravington, M. V. (2013). A Poisson–gamma model for analysis of ecological non-negative continuous data. *Environmental and Ecological Statistics*, 20(4):533–552.

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A., and Sibert, J. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27(2):233–249.

Fuentes, M., Chen, L., and Davis, J. M. (2008). A class of nonseparable and nonstationary spatial temporal covariance functions. *Environmetrics*, 19(5):487–507.

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. (2015a). Does non-stationary spatial data always require non-stationary random fields? *Spatial Statistics*, 14(C):505–531.

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. (2015b). Interpretable priors for hyperparameters for Gaussian random fields. *arXiv preprint arXiv:1503.00256*.

Garneau, D. E., Post, E., Boudreau, T., Keech, M., and Valkenburg, P. (2007). Spatio-temporal patterns of predation among three sympatric predators in a single-prey system. *Wildlife Biology*, 13(2):186–194.

Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M. (2010). *Handbook of spatial statistics*. CRC Press, Boca Raton.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association*, 97(458):590–600.

Gregory, R. D. and van Strien, A. (2010). Wild bird indicators: using composite population trends of birds as measures of environmental health. *Ornithological Science*, 9(1):3–22.

Griewank, A. and Walther, A. (2008). *Evaluating derivatives: principles and techniques of algorithmic differentiation*. Society for Industrial and Applied Mathematics, Philadelphia.

Held, L., Schrödle, B., and Rue, H. (2010). Posterior and cross-validatory predictive checks: a comparison of MCMC and INLA. In *Statistical modelling and regression structures*, pages 91–110. Springer, New York.

Ho, L. P. and Stoyan, D. (2008). Modelling marked point patterns by intensity-marked Cox processes. *Statistics & Probability Letters*, 78(10):1194–1199.

Högmander, H. and Särkkä, A. (1999). Multitype spatial point patterns with hierarchical interactions. *Biometrics*, 55(4):1051–1058.

Hu, X., Simpson, D., Lindgren, F., and Rue, H. (2013). Multivariate gaussian random fields using systems of stochastic partial differential equations. *arXiv preprint arXiv:1307.1379*.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. John Wiley & Sons, Chichester.

Illian, J. B. and Burslem, D. R. (2015). Improving the useability of spatial point process methodology: an interdisciplinary dialogue between statistics and ecology. Technical report, Trondheim University & University of St Andrews.

Illian, J. B., Martino, S., Sørbye, S. H., Gallego-Fernández, J. B., Zunzunegui, M., Esquivias, M. P., and Travis, J. M. (2013). Fitting complex ecological point process models with integrated nested Laplace approximation. *Methods in Ecology and Evolution*, 4(4):305–315.

Illian, J. B., Møller, J., and Waagepetersen, R. P. (2009). Hierarchical spatial point process analysis for a plant community with high biodiversity. *Environmental and Ecological Statistics*, 16(3):389–405.

Illian, J. B., Sørbye, S., Rue, H., and Hendrichsen, D. (2012a). Using INLA to fit a complex point process model with temporally varying effects–a case study. *Journal of Environmental Statistics*, 3(7).

Illian, J. B., Sørbye, S. H., and Rue, H. (2012b). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *The Annals of Applied Statistics*, 6(4):1499–1530.

Ingebrigtsen, R., Lindgren, F., and Steinsland, I. (2014). Spatial models with explanatory variables in the dependence structure. *Spatial Statistics*, 8(1):20–38.

Jackson, D., Vandermeer, J., Perfecto, I., and Philpott, S. M. (2014). Population responses to environmental change in a tropical ant: the interaction of spatial and temporal dynamics. *PloS one*, 9(5):e97809.

Jones-Todd, C. M., Caie, P., Illian, J., Stevenson, B. C., Savage, A., Harrison, D., and Bown, J. (In submission). Identifying unusual structures inherent in point pattern data and its application in predicting cancer patient survival.

Kauffmann, G. and Fairall, A. (1991). Voids in the distribution of galaxies: an assessment of their significance and derivation of a void spectrum. *Monthly Notices of the Royal Astronomical Society*, 248(2):313–324.

Kendall, W. S., Liang, F., and Wang, J.-S. (2005). *Markov chain Monte Carlo: innovations and applications*. World Scientific, Singapore.

Kim, H., Sun, D., and Tsutakawa, R. K. (2001). A bivariate Bayes method for improving the estimates of mortality rates with a twofold conditional autoregressive model. *Journal of the American Statistical Association*, 96(456):1056–1521.

Knorr-Held, L. and Best, N. G. (2001). A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164(1):73–85.

Kopeckỳ, J. and Mrkvička, T. (2016). On the Bayesian estimation for the stationary Neyman-Scott point processes. *Applications of Mathematics*, 61(4):503–514.

Krainski, E. T., Lindgren, F., Simpson, D., and Rue, H. (2016). The R-INLA tutorial on SPDE models. Technical report, Norwegian University of Science and Technology.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*, 70(5):1–21.

Langford, I. H., Leyland, A. H., Rasbash, J., and Goldstein, H. (1999). Multilevel modelling of the geographical distributions of diseases. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48(2):253–268.

Li, S. (2011). Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics and Statistics*, 4(1):66–70.

Lindgren, F. (2012). Continuous domain spatial models in `r-inla`. *The ISBA Bulletin*, 19(4):14–20.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Luchnikov, V., Medvedev, N., Oger, L., and Troadec, J.-P. (1999). Voronoi-Delaunay analysis of voids in systems of nonspherical particles. *Physical Review E*, 59(6):7205–7212.

Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. CRC press, Boca Raton.

Marchette, D. J. (2005). *Random graphs for statistical pattern recognition*. John Wiley & Sons, Hoboken.

Martins, T. G., Simpson, D., Lindgren, F., and Rue, H. (2013). Bayesian computing with INLA: new features. *Computational Statistics & Data Analysis*, 67(1):68–83.

Martins, T. G., Simpson, D. P., Riebler, A., Rue, H., and Sørbye, S. H. (2014). Penalising model component complexity: a principled, practical approach to constructing priors. *arXiv preprint arXiv:1403.4630*.

Matérn, B. (1986). *Spatial variation*. Springer-Verlag, Berlin.

Matheron, G. (1976). *A simple substitute for conditional expectation: the disjunctive kriging*. Springer, Netherlands.

Mattfeldt, T. and Fleischer, F. (2014). Characterization of squamous cell carcinomas of the head and neck using methods of spatial statistics. *Journal of Microscopy*, 256(1):46–60.

Menezes, R., Garcia-Soidán, P., and Febrero-Bande, M. (2005). A comparison of approaches for valid variogram achievement. *Computational statistics*, 20(4):623–642.

Møller, J. and Díaz-Avalos, C. (2010). Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires. *Scandinavian Journal of Statistics*, 37(1):2–25.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25(3):451–482.

Moller, J. and Waagepetersen, R. P. (2004). *Statistical inference and simulation for spatial point processes*. Chapman and Hall, London.

Møller, J. and Waagepetersen, R. P. (2007). Modern statistics for spatial point processes. *Scandinavian Journal of Statistics*, 34(4):643–684.

Myllymäki, M. and Penttinen, A. (2009). Conditionally heteroscedastic intensity-dependent marking of log Gaussian Cox processes. *Statistica Neerlandica*, 63(4):450–473.

Newson, S. E., Rexstad, E. A., Baillie, S. R., Buckland, S. T., and Aebischer, N. J. (2010). Population change of avian predators and grey squirrels in england: is there evidence for an impact on avian prey populations? *Journal of Applied Ecology*, 47(2):244–252.

Newton, I. (1986). *The Sparrowhawk*. Poyser, London.

Neyman, J. and Scott, E. L. (1958). Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 20(1):1–43.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27.

Priebe, C. E., DeVinney, J. G., and Marchette, D. J. (2001). On the distribution of the domination number for random class cover catch digraphs. *Statistics & Probability Letters*, 55(3):239–246.

Python, A., Illian, J., Jones-Todd, C. M., and Blángiardo, M. (2016). Explaining the lethality of Boko Harams terrorist attacks in Nigeria, 2009-2014: a hierarchical Bayesian approach. In *Conference proceedings: Bayesian Young Statisticians Meeting*.

Rajala, T. (2010). `spatgraphs`: Proximity description of spatial point patterns using graphs. Technical report, Department of Mathematics and Statistics, University of Jyvaskylà, Finland.

Rathbun, S. L. (1996). Estimation of Poisson intensity using partially observed concomitant variables. *Biometrics*, 52(1):226–242.

Richardson, S., Abellan, J. J., and Best, N. (2006). Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in yorkshire (uk). *Statistical Methods in Medical Research*, 15(4):385–407.

Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 39(2):172–212.

Robinson, R. A., Siriwardena, G. M., and Crick, H. Q. P. (2005). Size and trends of the house sparrow *Passer domesticus* population in Great Britain. *Ibis*, 147(3):552–562.

Rozanov, Y. A. (1982). *Markov random fields*. Springer, New York.

Rue, H. and Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of Statistical Planning and Inference*, 137(10):3177–3192.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Šálek, M. and Schröpfer, L. (2008). Population decline of the little owl (*Athene noctura scop*) in the Czech Republic. *Polish Journal Ecology*, 56(3):527–534.

Särkkä, S. and Hartikainen, J. (2012). Infinite-dimensional Kalman filtering approach to spatio-temporal Gaussian process regression. In *Conference proceedings: AISTATS*, volume 22, pages 993–1001.

Schoenberg, F. P. (2003). Multidimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association*, 98(464):789–795.

Schrödle, B. and Held, L. (2011). Spatio-temporal disease mapping using INLA. *Environmetrics*, 22(6):725–734.

Schweiger, O., Heikkinen, R. K., Harpke, A., Hickler, T., Klotz, S., Kudrna, O., Kühn, I., Pöyry, J., and Settele, J. (2012). Increasing range mismatching of interacting species under global change is related to their ecological characteristics. *Global Ecology and Biogeography*, 21(1):88–99.

Simpson, D. P., Martins, T. G., Riebler, A., Fuglstad, G.-A., Rue, H., and Sørbye, S. H. (2014). Penalising model component complexity: a principled, practical approach to constructing priors. *arXiv preprint arXiv:1403.4630*.

Šipkovskỳ, I. (2012). How the little owls successfully nested by us or the last known breeding of little owl in the tree hollow in Slovakia. *Dravce a sovy*, 8(2):10–11.

Soranio-Redondo, A., Jones-Todd, C. M., Bearhop, S., Hilton, G. M., Lock, L., Stanbury, A., Votier, S. C., and Illian, J. B. (In submission). Estimating species distribution in dynamic populations using point process models: a case study in the Eurasian crane and perspectives in ecology.

Sørbye, S. H. and Rue, H. (2014). Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spatial Statistics*, 8(1):39–51.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Spiegelhalter, D. J., Thomas, A., Best, N. G., Gilks, W., and Lunn, D. (1996). BUGS: Bayesian inference using Gibbs sampling. *Version 0.5,(version ii) http://www. mrc-bsu. cam. ac. uk/bugs*, 19.

Stein, M. L. (2005). Space–time covariance functions. *Journal of the American Statistical Association*, 100(469):310–321.

Stevenson, B. (2015). *nspp: Estimation of parameters for Neyman-Scott point processes*. R package version 0.0.1.

Stevenson, B. C., Borchers, D. L., and Fewster, R. M. (In submission). Cluster capture-recapture to account for identification uncertainty on aerial surveys of animal populations.

Stoyan, D. (1988). Thinnings of point processes and their use in the statistical analysis of a settlement pattern with deserted tillages. *Statistics*, 19(1):45–56.

Stoyan, D. and Penttinen, A. (2000). Recent applications of point process methods in forestry statistics. *Statistical Science*, 15(1):61–78.

Swallow, B., Buckland, S. T., King, R., and Toms, M. P. (2016). Bayesian hierarchical modelling of continuous non-negative longitudinal data with a spike at zero: an application to a study of birds visiting gardens in winter. *Biometrical Journal*, 58(2):357.

Tanaka, U., Ogata, Y., and Stoyan, D. (2008). Parameter estimation and model selection for neyman-scott point processes. *Biometrical Journal*, 50(1):43–57.

Taylor, B. M. and Diggle, P. J. (4). INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. *Journal of Statistical Computation and Simulation*, 84(10):2266–2284.

Thomson, D. L., Green, R. E., Gregory, R. D., and Baillie, S. R. (1998). The widespread declines of songbirds in rural Britain do not correlate with the spread of their avian predators. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1410):2057–2062.

Tu, S.-J. and Fischbach, E. (2002). Random distance distribution for spherical objects: general theory and applications to physics. *Journal of Physics A: Mathematical and General*, 35(31):6557.

Van Nieuwenhuyse, D., Génot, J.-C., and Johnson, D. H. (2008). *The little owl–Conservation, ecology and behavior of Athene noctua*. Cambridge University Press, Cambridge.

Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, 63(1):252–258.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41(3/4):434–449.

Wiegand, T., Gunatilleke, S., Gunatilleke, N., and Okuda, T. (2007). Analyzing the spatial structure of a Sri Lankan tree species with multiple scales of clustering. *Ecology*, 88(12):3088–3102.

Wöhlke, A., Philipp, U., Bock, P., Beineke, A., Lichtner, P., Meitinger, T., and Distl, O. (2011). A one base pair deletion in the canine atp13a2 gene causes exon skipping and late-onset neuronal ceroid lipofuscinosis in the Tibetan terrier. *PLoS Genet*, 7(10):e1002304.

Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466(7310):1102–1104.

Zammit-Mangion, A., Dewar, M., Kadirkamanathan, V., and Sanguinetti, G. (2012). Point process modelling of the Afghan war diary. *Proceedings of the National Academy of Sciences*, 109(31):12414–12419.

Zeldovich, I. B., Einasto, J., and Shandarin, S. (1982). Giant voids in the universe. *Nature*, 300(5891):407–413.

Acronyms

AD	Automatic Differentiation
BTO	British Trust for Ornithology
CCCD	Class Cover Catch Digraphs
CSR	complete spatial randomness
FEM	finite element methodoogy
GBFS	Garden Bird Feeding Survey
GMRF	Gaussian Markov Random Field
GRF	Gaussian Random Field
GTD	Global Terrorism Database
INLA	Integrated Nested Laplace Approximation
LGCP	log-Gaussian Cox process
LMC	linear models of coregionalisation
NSPP	Neyman-Scott point process
pair c.f	pair correlation function
Palm i.f	Palm intensity function
point pattern	point pattern
SPDE	stochastic partial differential equation
TMB	Template Model Builder

Mathematical notation

K_ν	modified Bessel function
$\mathbb{R}^d$	d dimensional space
$H(\cdot)$	empty space function (spherical contact distribution function)
$\stackrel{\mathrm{d}}{=}$	equally distributed
$\langle \cdot, \cdot \rangle$	inner product of two vectors
λ	point process intensity
Δ	Laplacian
$G(\cdot)$	nearest neighbour distribution function
$\setminus$	not contatining
Ω	some set, or loosley some spatial domain
o	the origin
$g(\cdot)$	pair correlation function
$\lambda_p(\cdot)$	Palm intensity function
$K(\cdot)$	Ripley's K-function
$N(\cdot)$	spatial point process
$S(\mathbf{w}_\mathbf{x}, w_t)$	spectral density
$b(x, r)$	sphere of radius r centred at some location x
$\boldsymbol{\theta}$	vector of parameters
$\mathcal{W}$	bounded observation window of a $N(\cdot)$

Species list

Accipiter nisus	Eurasian sparrowhawk
Athene noctua	little owl
Azteca sericeasur	tree-nesting ant
Coccus viridis	soft scale insect
Grus grus	crane
Passer domesticus	house sparrow
Rattus rattus	black rat or ship rat
Streptopelia decaocto	collared dove